Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption

H. X. Zhao, F. Magoules

Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption

Hai Xiang ZHAO, Frédéric MAGOULÈS
Ecole Centrale Paris, Applied Mathematics and Systems Laboratory, Grande Voie des Vignes, 92295 Châtenay-Malabry Cedex, France
Email: zhaohaix@ecp.fr, frederic.magoules@hotmail.com

ABSTRACT
Analyzing the energy performance in a building is an important task in energy conservation. To accurately predict the energy consumption is difficult in practice since the building is a complex system with many parameters involved. To obtain enough historical data of energy uses and to find out an approach to analyze them become mandatory. In this paper, we propose a simulation method with the aim of obtaining energy data for multiple buildings. Support vector machines method with Gaussian kernel is applied to obtain the prediction model. For the first time, a parallel implementation of support vector machines is used to accelerate the model training process. Our experimental results show very good performance of this approach, paving the way for further applications of support vector machines method on large energy consumption datasets.

Keywords: Energy consumption, building, prediction, support vector machines, parallel computing, grid computing

1 INTRODUCTION

Nowadays, energy conservation is a critical task. It is involved in almost every aspect of energy flow including exploitation, delivery and end-use. Predicting the energy consumption of a building is an important approach in energy conservation which benefits both individual and society. With the basic knowledge of the energy performance of a building, it is possible to more wisely design a new building.

However, it is difficult to realize in practice. The first reason lies on the fact that a building is a rather complex system with a great number of influencing factors involved. Taking the heating load for example, it is influenced by the outside weather conditions, the structures, the thermal behavior of the materials, the inner facilities and the behavior of the occupants. Another important reason is that there is not enough data available because of the difficulties in gathering valuable data from the real world [1]. The third reason lies in the difficulty of analyzing the historical data due to the lack of methods, especially regarding the time consuming aspects when dealing with large datasets.

In order to achieve accurate predictions, it is necessary to collect enough historical data and to find out a practical way to analyze these data. This paper generates data of multiple buildings from simulations with the EnergyPlus [2] software. A well known statistical learning theory, the support vector machines (SVMs) method [3], together with the Gaussian kernel (RBF) are applied to obtain the prediction model. A parallel implementation of SVMs is used for the first time to accelerate the model training process on large datasets. Our experimental results show that SVMs method has a good performance in predicting unknown buildings, and that the parallel implementation makes it possible to analyze large energy consumption datasets in practice. Section 2 of this paper briefly introduces the related work, and Section 3 introduces the learning theory and its parallel implementation. How to collect the data is introduced in Section 4. Description of the experiments and analysis of the results are presented in Section 5. Finally, the conclusions are presented in Section 6.

2 RELATED WORK

In recent years, the analysis of building energy performance is wildly studied. A large number of software have been developed for evaluating energy efficiency and sustainability of buildings [2, 4], such as DOE-2, AkWarm, Antherm, Apache, etc. Most of them are simulation tools which calculate, on an hourly or monthly basis, the thermal transfer or energy variation of buildings and facilities from an engineering point of view.

At the same time, researchers have developed several analyzing models to predict the energy performance based on historical data. The regression model proposed in [5–7] is well suited for long period energy predictions and is easy to develop. The time-series analysis [8] and the Fourier series model [9] were proposed for analyzing historical time series data. The neural network related methods are the most frequently applied approaches [10–14]. Some of these methods have been thoroughly tested and thus can be used as benchmark calculations for other prediction methods.

Support vector machines (SVM) are a set of methods for classification and regression. It is close to multi-layer neural network and has good generalization abilities in solving non-linear problems [3, 15]. Despite it is widely used in industries, only few work has been done for applying it to analyze building energy behaviors [16–18].

Dong et al. [16] first applied SVMs to predict the monthly electricity consumption of four buildings in tropical region. Three years' data was trained and the derived model was applied on one year's data to predict the landlord utility in that year. The results showed good performances of SVMs on this problem.

Lai et al. [17] applied robust regression model on one year's data of a building. The dataset was recorded on a daily basis with electricity consumption and climate variation involved. In these experiments, the authors trained the model on one year's data and then applied it to three month's data to test the predication ability of this derived model. The authors also trained the models on each daily basis dataset and then compared the obtained models to verify the stability of this approach. In addition, perturbation to certain data has been added and detection of the perturbation by examining the change of the contributing weights has been investigated.

Li et al. [18] used SVMs to predict the hourly cooling load of an office building. The performance of the support vector regression was compared with the conventional back-propagation neural networks and it was proved to be better than the traditional solution.

The above three works have shown that SVMs can provide good performances in predicting hourly and monthly building energy. However, there are several shortages in these studies. Firstly, all the experiments were performed on a small number of buildings which means that we do not know the ability of SVMs in predicting the energy performance in a completely new building for instance. Secondly, the datasets for model training were not large enough, which could lead to a certain degree of limitations of these models. For instance, the relation between the energy consumption and a limited number of features such as weather conditions are considered, and the time duration is not long enough.

3 THEORY OF SUPPORT VECTOR MACHINES

3.1 Support vector regression

Support vector machine is a supervised learning method which aims at finding a decision function to represent the relationship between the features and the target. This function is also called a model or a pattern. The features are also called the variables. According to different types of targets, SVM is classified into two types: one is the classification in which the target has only two values, e.g. $\{0, 1\}$, and the other type is the regression in which the target has continuous real value.

Let vector x_i represents the ith sample of the features and y_i represent the corresponding target value, therefore, all of the samples can be represented as:

$$(x_1, y_1), (x_2, y_2), ...(x_l, y_l)$$

where $x_i \in \mathbb{R}^n$ and means that there are n features, $y_i \in \mathbb{R}$ and l is the number of samples. We are aiming at finding a decision function of the form:

$$f(x) = \omega x + b \tag{1}$$

where $f(x_i)$ is the estimation of the corresponding y_i. If such a function is found, we can use it to predict the unknown y_i with the new input x_i. This is defined as the prediction ability of the derived model. The process to find (1) is called model training. In the above function, ω and b are coefficients. The main task now is to find out the proper values of ω and b. We approximate them by minimizing the empirical risk with respect to the loss function:

$$L(y - f(x)) = \sum_{i=1}^{l} |y_i - f(x_i)| \tag{2}$$

In this paper, the target is the total energy consumption in a building–which is a real continuous variable. Thus we only use support vector regression (SVR) to estimate the target, and we talk about SVR. To make our estimation robust, we use an ε-insensitive loss function instead of the above quadratic loss function.

$$L(y - f(x)) = \begin{cases} 0 & \text{if } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases}$$

This loss function corresponds to an ε-tube around the measured target values. We assume that there is no deviation of the predicted values from the measured ones if they lie inside the tube. To find out the vector ω, the problem is equivalent to the following quadratic optimization problem after introducing two slack variables ξ_i and ξ_i^* , $i = 1, 2, ..., l$, and where we minimize

$$\frac{1}{2}\|\omega\|^2 + C(\sum_{i=1}^{l} \xi_i^* + \sum_{i=1}^{l} \xi_i) \tag{3}$$

subject to the constraints

$$y_i - f(x_i) \leq \varepsilon + \xi_i^*$$

$$f(x_i) - y_i \leq \varepsilon + \xi_i$$

$$\xi_i^*, \xi_i \geq 0, \qquad i = 1, 2, ..., l$$

where C is a regularizing constant, which determines the trade off between the capacity of $f(x)$ and the number of points outside the ε-tube. To find the saddle point of the function (3) under the previous inequalities constraints, one can turn to the Lagrange function by introducing four Lagrange multipliers, α^*, α, γ^*, γ. The Lagrange function becomes:

$$\begin{aligned} L(\omega, b, \xi^*, \xi, \alpha^*, \alpha, \gamma^*, \gamma) = {} & \frac{1}{2}\|\omega\|^2 \\ & + C(\sum_{i=1}^{l} \xi_i^* + \sum_{i=1}^{l} \xi_i) \\ & - \sum_{i=1}^{l} \alpha_i[y_i - (\omega x_i) - b + \varepsilon + \xi_i] - \alpha_i^*[y_i - (\omega x_i) - b + \varepsilon + \xi_i^*] \\ & - \sum_{i=1} l(\gamma_i^* \xi_i^* + \gamma \xi) \end{aligned} \tag{4}$$

The four Lagrange multipliers satisfy the constraints $\alpha^* \geq 0, \alpha \geq 0, \gamma^* \geq 0$ and $\gamma \geq 0, i = 1, 2, ..., l$. If the relations

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi^*} = \frac{\partial L}{\partial \xi}$$

occur, we get the following conditions,

$$\omega = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) x_i$$

$$\sum_{i=1}^{l} \alpha_i^* = \sum_{i=1}^{l} \alpha_i \tag{5}$$

$$0 \leq \alpha_i^*, \alpha_i \leq C \tag{6}$$

$$C = \alpha_i^* + \gamma_i^* = \alpha_i + \gamma_i, \qquad i = 1, 2, ..., l \tag{7}$$

Putting them back into the above Lagrange function, we obtain the solution of the optimization problem which is equal to the maximum of the function (4) with respect to the Lagrange multipliers. The next step is to find α_i^* and α_i in order to maximize the following function:

$$W(\alpha_i^*, \alpha_i) = \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) - \varepsilon \sum_{i=1}^{l} (\alpha_i^* + \alpha_i) - \frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) \tag{8}$$

under the constraints (5) and (6), where $x_i \cdot x_j$ stands for the dot product of two vectors x_i and x_j. Normally, only some of the samples satisfy the property of $\alpha_i^* - \alpha_i \neq 0$, which are called support vectors (SVs). In fact, only these samples lying outside the ε-tube will contribute to determine the decision function Eq. (1). In practice, it is difficult to find out a linear function $f(x)$ for real problems with a large dataset. In such cases, one often maps the 1-dimensional problem into a higher dimensional feature space where it is easier to find a linear function similar to the previous decision function. Fortunately, it is not necessary to express the mapping explicitly. Instead, the final form $f(x_i) = \varphi(x_i) + b$ is good enough. Thus, the last term $x_i \cdot x_j$ of (8) is changed into $\varphi(x_i) \cdot \varphi(y_i)$ which is called the kernel function and the term $K(x_i \cdot y_i)$ is used to replace it. Then, combining with Eq. (4), the decision function goes to:

$$f(x) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) K(x_i \cdot x) + b \tag{9}$$

where α_i^* and α_i are determined by maximizing the quadratic function:

$$W(\alpha_i^*, \alpha_i) = \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) - \varepsilon \sum_{i=1}^{l} (\alpha_i^* + \alpha_i) - \frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i \cdot x_j) \tag{10}$$

under the constraints (5) and (6).

There are four frequently used kernel functions in SVMs, which are the linear function, the polynomial function, the radial basis function (RBF) and the sigmod function. They represent different decision shapes in the feature space. In this paper, we chose RBF in the training process because it has been tested to be proper for some industrial applications [16]. RBF is also called Gaussian kernel, with the form $K(x_i, x_j) = exp(-\gamma\|x_i - x_j\|^2)$ where $\gamma > 0$ is the kernel parameter.

3.2 Solving the quadratic problem

When solving Eq. (10) via standard quadratic problem (QP) techniques, there would be a matrix in the memory whose number of elements is equal to the square of the number of samples [19]. If the trained data is very large, it would become difficult to load the whole matrix into the computer memory. Furthermore, there are a significant number of numerical problems when solving such a function. Therefore, sequential minimal optimization (SMO) method was proposed and has been proved to be a better alternative to handle these two difficulties when dealing with large datasets. The main idea of SMO is to divide the entire optimization problem into sub-problems and to solve each sub-problem step by step, until finally conquer the whole problem. In each step, two Lagrange multipliers are optimized analytically until the Karush-Kuhn-Tucher (KKT) conditions are satisfied. Which two Lagrange multipliers are selected in current step is decided by an heuristics algorithm. The parameter b is re-computed in each step. More details on the SMO method could be found in [15, 19].

3.3 Parallel implementation

SVM training becomes a time consuming process when applied to large datasets. Brugger [15] integrated SMO method to solve the quadratic problem in parallel. By profiling the performance of the optimization process in different datasets, the author found that the most time-consuming part is located in the kernel evaluation. Therefore, he parallelized the kernel evaluations and gradient updates, combined with inner sequential QP solver and distributed storage of kernel rows, achieved linear speedup for regression in his test. The tool Pisvm was developed using this idea.

In this paper, we trained our model on multiple buildings datasets with the Pisvm tool for the purpose of accelerating the training process in high burden conditions.

3.4 Performance evaluation

Two performance evaluation methods were applied in this work. One is the mean squared error (MSE) which gives the average deviation of the predicted values to the measured one. The lower the MSE, the better the performance of the prediction. The other method is the squared correlation coefficient (SCC) which lies between $[0, 1]$ and gives the ratio of successfully predicted number of target values on the total number of target values, i.e. how certain the predicted values are compared to the measured one. The higher the SCC, the stronger the evaluating ability.

4 DATA PREPARATION

Buildings for office use located in Paris-Orly were simulated via EnergyPlus. We chose this tool because, as a succession of the well-known building energy simulation software BLAST and DOE-2.1E, it is well tested and comprehensive in calculating energy performance of complex systems. In our work, we generated the data of a single building in heating season initially. Then, by modifying some alterable parameters, we generated the data for multiple houses.

Before running the simulation, one has to create an input file (`idf` file) to give the parameters of a building for EnergyPlus. The input parameters include the weather conditions, the building structures, the inside occupants' behaviors, the schedule of light using, etc. The most important parameters are shown in Table 1 and the materials of the surfaces are given in Table 2. The description of these materials can be found in the documents of EnergyPlus [2].

Table 1: Input parameters of a single building (in metric units).

Parameters	Values
Location	Paris-orly, City
Duration	From Nov 1 to Mar 31
Building Shape	Rectangle
Structure	Length:11 Width:10 Ceiling Height:4 North axis: 10^o
Fenestration surface	$14m^2$ for each wall
Thermal Zones	1
People	14
Air infiltration	0.0348 m^3/s
Heating type	District heating
Cooling type	HVAC windowAirCondioner
Other facilities	Light, Water heater

Table 2: Building materials used for the simulation.

Structures	Material's name	Thickness(m)	Conductivity(W/mK)
Wall	1IN Stucco	0.0253	0.6918
	8IN Concrete HW	0.2033	1.7296
	Wall Insulation	0.0679	0.0432
Ground	MAT-CC05 8 HW CONCRETE	0.2032	1.311
Roof	Roof Membrane	0.0095	0.16
	Roof Insulation	0.1673	0.049
	Metal Decking	0.0015	45.006
Windows	Theoretical Glass [117]	0.003	0.0185

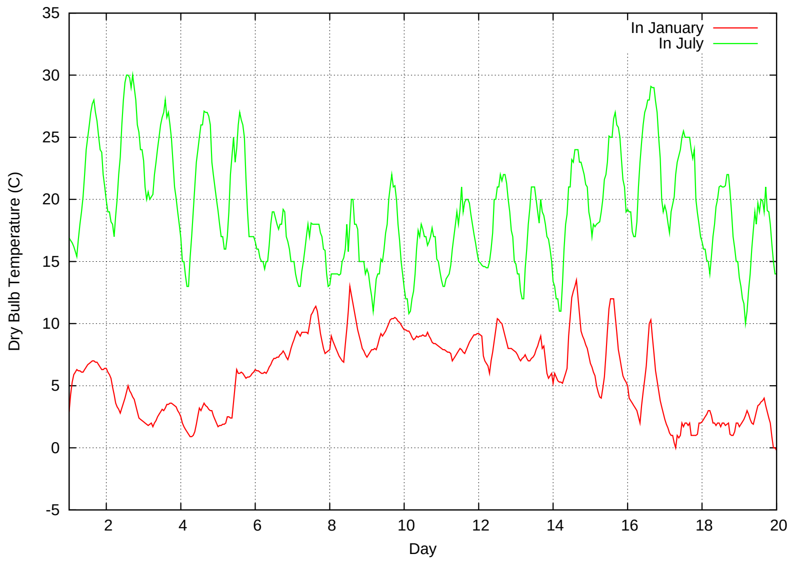

Figure 1: Dry bulb temperature in the first day of each month.

We supposed that there was only one floor and one room in this building. Because it was in heating season, the energy consumed by this building mainly comes from three sources which are the district heating used to keep the inside space warm, the electricity lights used mostly in working days, and the hot water for office use. For the walls in each orientation, there are several construction layers for thermal reasons. That explains why there are three materials for the walls as indicated in Table 2. The same occur for the roof. Besides these parameters, we used hourly recorded weather data in Paris-Orly along with other important input parameters. The weather data includes dry bulb air temperature, relative humidity, global horizontal radiation and ground temperature. To have a glance of these data, the dry bulb air temperatures of the first 20 days in January and July are plotted in Figure 1, with the relative humidity in the same days plotted in Figure 2.

In this simulation, the output was hourly damped. There were several output files in EnergyPlus, where we extracted useful data, mainly from the `eso` file. As the analyzing step required, we have to reformat this data into the form required by the analyzing tools. We take district heating demand or total electricity consumption as the target. Meanwhile, we take 25 variables as the features for a single building, which are day type indications if the current day is holiday or not, weather conditions, zone mean air temperatures, infiltration volume, heat gain through each window, heat gain through lights and people, zone internal total heat gain.

In order to generate the data for multiple buildings, we developed an interface to automatically control the simulation process since there was no available user interface for EnergyPlus for this purpose. In our approach, the input file

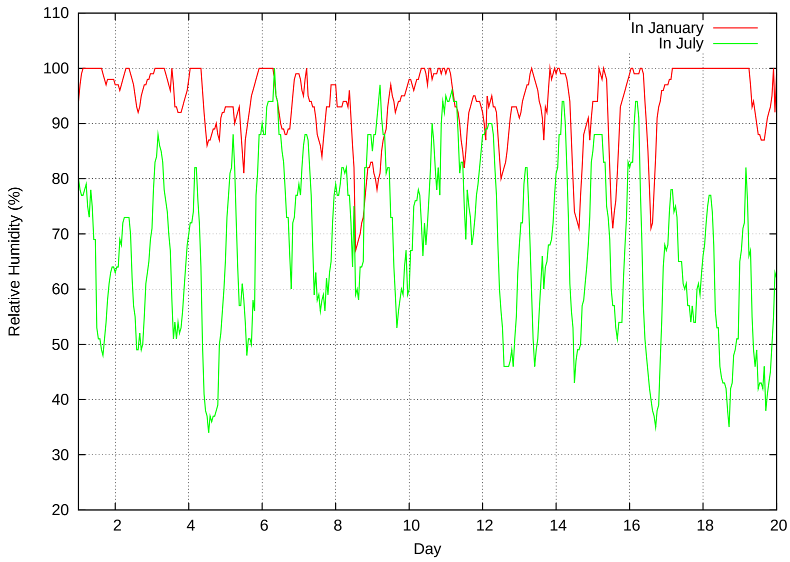

Figure 2: Relative humidity in the first day of each month.

was divided into two parts. The first part is called the alterable part, containing the parameters which would be different for each building. The values of these parameters were obtained by stochastic methods. The second part is called stable part where the parameters are always the same for each building. After updating the alterable part for a new building, we combined this part with the stable part to create the final input file. To analyze multiple buildings, we have to put all the generated output data of each building into a single output file, which is named as `output.txt`. Then, the program constantly updates the alterable part to create the data for the next building. The above steps of generating the multiple buildings are shown in Figure 3.

5 EXPERIMENTS AND RESULTS

In supervised learning theory, the experiments can be roughly divided into two steps, training and predicting. Accordingly, the data is divided into two parts, one is for training, called a training set, and the other is for predicting, called a testing set. A decision model is obtained in the training step on the training set to indicate the dependence of the target on the features. In the predicting step, the trained model is applied on the testing set to predict the target values with regard to new features. By comparing the predicted target with the measured one, it is possible to evaluate the prediction performance of the model.

Before training the data by SVR, we need to scale the values linearly into a small range in order to avoid numerical problems in the training. Here we chose the range $[0, 1]$. Likewise, the testing values should also be scaled with the same

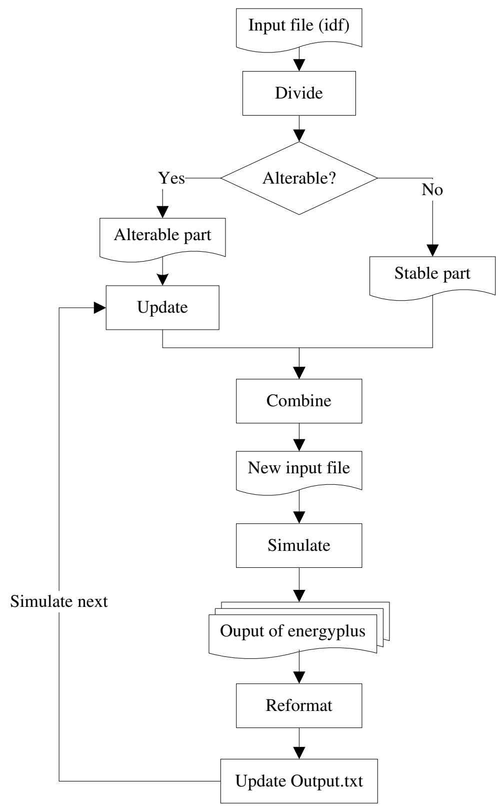

Figure 3: Flow chart of generating energy consumption data of multiple buildings.

scaling function. Therefore, it is possible for the scaled test values to be in a different range from that of the training values.

To choose the optimal values for the SVM parameters is really crucial in practice. The best parameters should have the ability to well predict on unknown data without the over-fitting problem. In our experiments, we selected SVR with RBF kernel to train our model. The parameters needed were C, γ, ε. The estimation of γ was solved by $\gamma = \sum_{i,j=1}^{l}(\left\|x_i - x_j\right\|^2)$ as proposed in [20] and [15]. The SVR parameters C and ε were solved by 5-fold cross validation on randomly selected 3000 samples from the training set. The above steps necessary for our experiments are shown in Figure 4.

Three experiments have been performed on different types of datasets. The first one is to analyze energy consumption data for a single building in order to test the prediction performance of SVMs method on building energy. The second one is to analyze multiple buildings with detailed building structures involved. The third one is to test the performance of parallel SVMs on a large dataset from multiple buildings. The first and the second experiments were done in Libsvm [21] which is a widely used sequential implementation of SVMs.

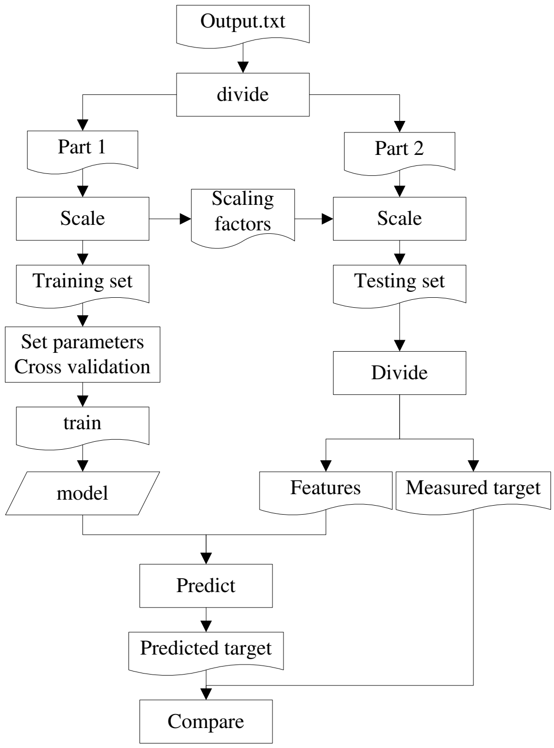

Figure 4: Flow chart of a learning process.

The third one was performed in the parallel implementation tool — Pisvm. The experimental environment is a cluster composed by two homogeneous workstations linked with Gigabit Ethernet. Each PC has $8 * 2.5$GHz CPU, 1333MHz FSB and 4G memory. The operating system is Linux (kernel version 2.6).

In the first experiment, the district heating consumption data was gathered hourly from November 1st to March 31st. There were 3624 samples with 24 features in the final dataset. We took the samples of the last two days as testing set, and the rest of them for training use. The number of training samples was 3576 and the number of testing samples was 48. The parameters were set as $C = 16$, $\gamma = 0.7533$ and $\varepsilon = 0.01$. The result of the learning and predicting processes showed that the number of support vectors (SVs) was 2229, MSE was $2.3e-3$ and SCC was 0.927918. The measured and predicted targets are plotted in Figure 5 which shows very good prediction performance of the model.

Another important energy type consumed by buildings is electricity. We also trained a similar model to predict the electricity consumption through one year. To be different from the first experiment, this time we randomly selected 48 samples as testing set for the evaluation of the model. The features were the same as in the first experiment. The number of training samples was 8712. The parameters were set as $C = 16$, $\gamma = 0.3043$ and $\varepsilon = 0.01$. The result

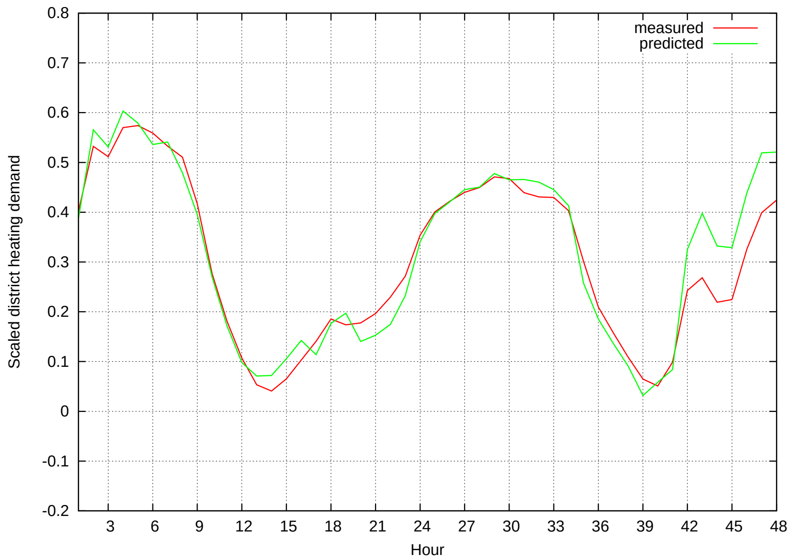

Figure 5: Measured and predicted district heating demand in heating season in the first experiment.

showed that, there were 2126 support vectors, MSE was $5.27e-4$, and SCC was 0.959905. The measured and predicted values are plotted in Figure 6. The model performs very well in this situation, except in the hours 18 and 28.

The above two learning processes are based on a single building energy consumption data. The evaluation of the model is to predict the unknown future in the same building. In practice, it is quite useful to predict how much energy would be used in a completely new building. Therefore, in the second experiment, we tried to learn a model based on the energy data according to the building structures involved. That is to say, we trained a model from the consumption behaviors of several buildings, then applied the model to predict the behavior of a different building. In this experiment, one hundred buildings were simulated in the heating season. They were in the same weather conditions but have different properties, such as different orientations, volumes, people densities and fenestration. We chose the data of the first 99 buildings as the training set and data of the last building as the testing set. The number of features was 28. The number of training samples was 358776, and the number of testing samples was 3624. The parameters were set as $C = 4$, $\gamma = 0.3179$ and $\varepsilon = 0.01$. In the training step, the number of SVs was 27501, while in the predicting step, MSE was $5.01e-5$ and SCC was 0.997639. The predicted and measured values on the first 100 samples in test dataset are plotted in Figure 7. The experimental results prove that SVR has a very good prediction performance in building energy consumption when building diversity is taken into account. It gives us the possibility to predict the energy performance of a building for designing as well as for retrofitting.

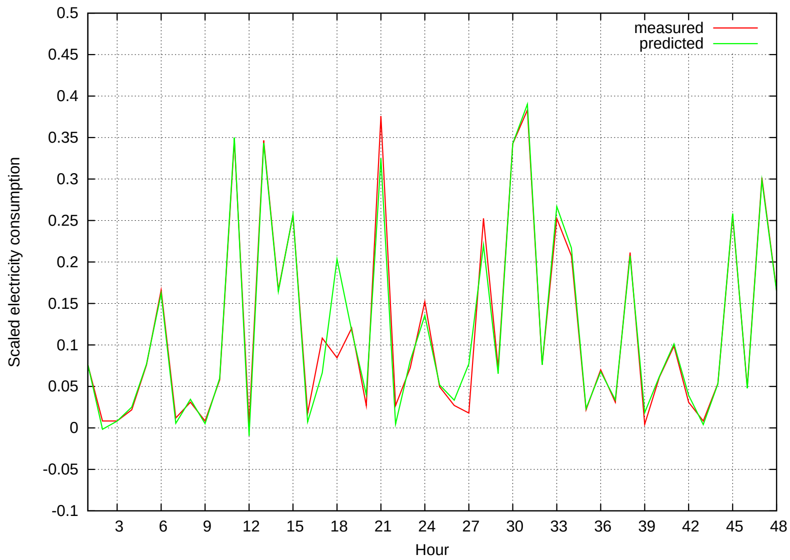

Figure 6: Measured and predicted electricity consumption through one year in the first experiment.

To increase the accuracy of the prediction, we have to collect energy consumption data as much as possible. But the problem is that the analyzing time increases quite fast while the data size is growing. In the second experiment, the most time-consuming process is the training step which needs 31.4 hours in our experimental environment. It becomes really long for analyzing 100 buildings. It is obvious that the time would go even longer if we calculate more buildings. To make the model more practical, it is necessary to reduce the learning time. As explained in Section 4, we turn to a parallel approach of SVMs in the next experiment. To test the speedup of this parallel solution, we did a set of experiments on 1, 2, 4, 6, 8 processors with the cache set as 256MB for each process. Each experiment was repeated three times and the average training time and speedup were calculated as shown in Figure 8. From the plotting, we can see that the speedups in 2 and 4 processors is close to linear speedup. But the performance enhancement is not obvious in 6 and 8 processors. One reason is that, in our testing environment with SMP structure inside the node, the distributed cache approach could not achieve its effect when the number of processes is big. Another reason is the low speed network connection between two nodes. However, from the varying trend in 2 and 4 processors, a better speedup can be expected on servers connected by higher speed devices such as infiniBand. The prediction performances, MSE and SCC, keep stable in each experiment. The number of SVs, MSE and SCC are compared with the sequential implementation as shown in Table 3.

Table 3 indicates that the result of parallel implementation is quite close to the sequential one, which means that the prediction performance of the parallel

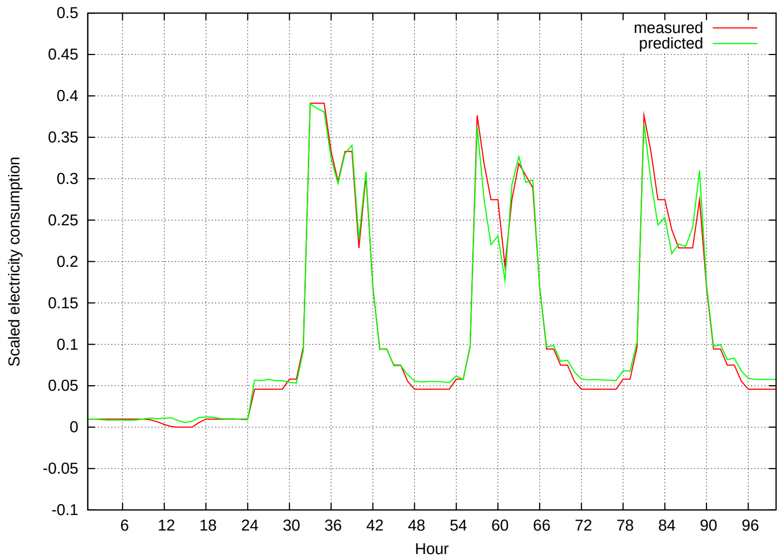

Figure 7: Measured and predicted electricity consumption in the second experiment.

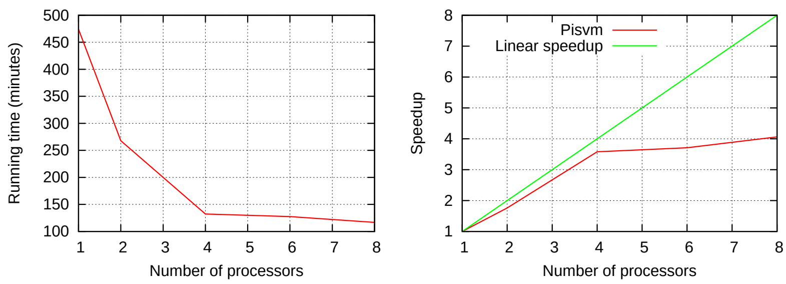

Figure 8: (a) Running time of the training process in parallel implementation of SVMs. (b) The speedup in comparison with linear speedup.

solution on large datasets is quite close to the sequential implementation on smaller datasets. The parallel solution can be applied to predict the building energy performances in more complex situations.

6 CONCLUSION

In this paper, a parallel implementation of support vector machines (SVM) method is applied for the first time to the prediction of energy consumption in multiple buildings based on large time series datasets. SVM has showed good performance in the prediction of building energy performance for some related work [16–18]. Since building energy problems are quite complex, we

Table 3: Comparison of parallel and sequential implementations.

Implementations	SVs	MSE	SCC
Sequential	27501	5.01097e-05	0.997639
Parallel	27382	5.08532e-05	0.997571

have to obtain historical data as much as possible for making our prediction more accurate. Consequently, the learning process of SVM would become very slow as the size of data increases.

This work introduced an approach of collecting abundant energy consumption data of multiple buildings by simulation. The prediction of energy performance in a completely new building was performed. A parallel approach of SVM is applied on those data to accelerate the training process. SVR with carefully chosen parameters showed a good performance in these experiments. Parallel SVM has a strong potential for analyzing extremely large energy data which would contribute to the accuracy of prediction in more complex situations in practice.

However, the evaluation of our work based on simulated buildings is not good enough. The better way is to record the real energy consumption data in various types of buildings. In addition, the speedup is not as good as linear one. To improve it, there are at least two possibilities, one is to take more consideration on selecting the features, the other is to improve the parallel solution of SVMs.

References

[1] L. Pérez-Lombard, J. Ortiz, and C. Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.

[2] EnergyPlus, 2008. Available online at: `http://www.EnergyPlus.gov` (Accessed July 2009).

[3] V.N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[4] M.S. Al-Homoud. Computer-aided building energy analysis techniques. *Building and Environment*, 36:421–433, 2001.

[5] M. Fels. Special issues devoted to measuring energy savings: the scorekeeping approach. *Energy and Buildings*, 9(1-2):5–18, 1986.

[6] D. Ruch and D.E. Claridge. A fourth parameter change-point model for predicting energy consumption in commercial buildings. *ASME Journal of Solar Energy Engineering*, 114(2):77–83, 1992.

[7] S. Katipamula, T.A. Reddy, and D.E. Claridge. Multivariate regression modeling. *ASME Journal of Solar Energy Engineering*, 120:177–184, 1998.

[8] A. Kimbara, S. Kurosu, R. Endo, K. Kamimura, T. Matsuba, and A. Yamada. On-line prediction for load profile of an air-conditioning system. *ASHRAE Transactions*, 101(2):198–207, 1995.

[9] A. Dhar, T.A. Reddy, and D.E. Claridge. Modeling hourly energy use in commercial buildings with fourier series functional form. *ASME Journal of Solar Energy Engineering*, 120:217–223, 1998.

[10] 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*, In Press, 2009.

[11] A. Azadeh, S.F. Ghaderi, and S. Sohrabkhani. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management*, 49(8):2272–2278, 2008.

[12] Alberto Hernandez Neto and Flvio Augusto Sanzovo Fiorelli. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, 40(12):2169–2176, 2008.

[13] 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.

[14] B.B. Ekici and U.T. Aksoy. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5):356–362, 2009.

[15] D. Brugger. Parallel support vector machines. In *VLSI-SoC07: Proceedings of the IFIP International Conference on Very Large Scale Integration of System on Chip*. Springer, 10 2007.

[16] 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.

[17] 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.

[18] Q. Li, Q.L. Meng, J.J. Cai, Y. Hiroshi, and M. Akashi. Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10):2249–2256, 2009.

[19] J.C. Platt. *Fast training of support vector machines using sequential minimal optimization.* MIT Press, Cambridge, MA, USA, 1999.

[20] I.W. Tsang, J.T. Kwok, and K.T. Lai. Core vector regression for very large regression problems. In *ICML'05: Proceedings of the 22nd International Conference on Machine Learning*, pages 912–919, New York, NY, USA, 2005. ACM.

[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 July 2009).