



HAL
open science

A parallel statistical learning approach to the prediction of building energy consumption based on large datasets

H. X. Zhao, F. Magoules

► **To cite this version:**

H. X. Zhao, F. Magoules. A parallel statistical learning approach to the prediction of building energy consumption based on large datasets. 6th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES), Oct 2009, Wuhan, Hubei, China. hal-00617926

HAL Id: hal-00617926

<https://hal.science/hal-00617926>

Submitted on 31 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Parallel Statistical Learning Approach to the Prediction of Building Energy Consumption Based on Large Datasets

Hai Xiang ZHAO, Frédéric MAGOULÈS
Applied Mathematics and Systems Laboratory, Ecole Centrale Paris,
Grande Voie des Vignes, 92295 Châtenay-Malabry Cedex, France
Email: zhaohaix@ecp.fr, frederic.magoules@hotmail.com Tel.: +33 (0)1 41 13 10 00

ABSTRACT

The prediction of future energy consumption of buildings based on historical performances is an important approach to achieve energy efficiency. A simulation method is here introduced to obtain sufficient clean historical consumption data to improve the accuracy of the prediction. The widely used statistical learning method, Support Vector Machines (SVMs), is then applied to train and to evaluate the prediction model. Due to the time-consuming problem of the training process, a parallel approach is applied to improve the speed of the training of large amounts of data when considering multiple buildings. The experimental results show very good performance of this model and of the parallel approach, allowing the application of Support Vector Machines on more complex problems of energy efficiency involving large datasets.

Keywords: Support Vector Machines (SVMs), Prediction, Model, Energy Efficiency, Parallel Computing.

1. INTRODUCTION

Building energy consumption plays an important role in the total energy consumption of end use. Energy efficiency in building management and retrofit is important for energy conservation which significantly benefits both individuals and society. For a specific building, the energy transformation is treated as a complex system with many factors involved, such as building construction and materials, ambient weather conditions, occupants' behaviors, inner facilities, etc. In fact, most of these factors vary irregularly from one building to another building and from time to time, making analysis of the energy performance extremely difficult in practice. Many works have been done in the past few decades in order to find out the empirical rules of energy performance. Some work aimed to simulate and evaluate the energy efficiency of buildings or facilities from the engineering point of view [1, 2]. A large number of tools were developed such as DOE-2, Apache, AkWarm, etc. Other works took the historical energy consumption behaviors into consideration, and tried to find out the rules from those recorded data and then applied these rules to predict the performance in unknown periods or conditions. Some models and methods were proposed for this kind of data mining approach, e.g. regression model [3, 4], time-series analysis [5], neural networks [6–10] and statistical learning method [11–13]. This last method is widely used in industry to analyze complex real-life problems due to the great generalization ability of Support Vector Machines (SVMs) in solving non-linear problems [14, 15].

Dong et al. [11] first applied SVMs to predict the monthly electricity consumption of four buildings in the tropical region. Three years' data was trained and the derived model was applied to predict the landlord utility in one year. Lai et al. [12] applied robust regression model on one year's data of one

building with electricity consumption and climate involved, then they applied the model to three month's data to test the prediction ability of their model. They also trained models on each daily basis datasets and then compared the obtained models to verify the stability of this method. In addition, they added some perturbations to a certain part of the data and tried to detect these perturbations by examining the change of contributing weights. Li et al. [13] compared SVMs with the conventional back-propagation neural network in predicting the hourly cooling load of an office building. SVMs method was proved to be better than the traditional solution. The above three work has shown good performances of SVMs method in predicting hourly and monthly building energy consumption. However, some problems remain in these analyses. Firstly, the predictions are based on the historical performance of the same building, which means that we do not know the ability of SVMs in predicting the energy performance in a completely new building. Secondly, only the dependence of energy requirements on a limited number of features, such as weather conditions, is considered. Obviously this limitation leads to inaccuracy in the previous models.

To improve the accuracy of the predicting model, this paper presents a simulation method to obtain enough energy consumption data for multiple buildings. Also, it applies SVMs to predict the performance of a new building with construction parameters involved. A parallel implementation is then applied to very large dataset in order to make SVMs more practical in solving complex problems. The plan of the paper is the following. In section 2, we briefly introduce the theory of SVMs method and its parallel approach. In section 3, we present the simulation method for obtaining energy consumption data of multiple buildings. The data analysis and numerical experiments are then presented in section 4. Finally, conclusions are given in section 5.

2. SUPPORT VECTOR REGRESSION METHOD

Support vector machines aim at finding a decision function (model) to represent the relationship between the features and the target. Let vector x_i indicates the i th sample of the features and y_i represents the corresponding target value. Therefore, all the samples can be represented as:

$$(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$$

where $x_i \in K^n$, $y_i \in K$ and l is the number of samples. If y is a continuous variable with real values, then this method is called a support vector regression (SVR). The decision function we are looking for can be expressed as:

$$f(x) = \omega x + b \quad (1)$$

where $f(x_i)$ is the estimation of the corresponding y_i and where ω and b are coefficients. If such a function is found, we can use it to estimate the unknown y_i with the new feature x_i (this is called the prediction ability of the extracted model). The

process to find this function is usually called model training with the main task to find out the proper values of ω and b . We approximate them by minimizing the empirical risk with respect to the following 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 ε -insensitive function supposes an assumption that there is no deviation of the estimated value to the measured one if their distance is less than ε . That means that we only consider the points outside the ε -tube around the measured values while we are optimizing the decision function. To find out the vector ω , the problem is equivalent to the following quadratic optimization problem with two new variables ξ_i and ξ_i^* ,

$i = 1, 2, \dots, l$, where we minimize

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

under the constraints

$$\begin{aligned} y_i - f(x_i) &\leq \varepsilon + \xi_i^* \\ f(x_i) - y_i &\leq \varepsilon + \xi_i \\ \xi_i^*, \xi_i &\geq 0, \quad i = 1, 2, \dots, l \end{aligned}$$

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 solve the problem of Eq. (2), we can use a Lagrange function by introducing four Lagrange multipliers, $\alpha^*, \alpha, \gamma^*, \gamma$.

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

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

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi_i^*} = \frac{\partial L}{\partial \xi_i} = 0$$

occur, we can get the following conditions,

$$\omega = \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i \quad (4)$$

$$\sum_{i=1}^l \alpha_i^* = \sum_{i=1}^l \alpha_i \quad (5)$$

$$0 \leq \alpha_i^*, \alpha_i \leq C \quad (6)$$

$$C = \alpha_i^* + \gamma_i^* = \alpha_i + \gamma_i, \quad i = 1, 2, \dots, l \quad (7)$$

Putting them back into the Lagrange function, we can obtain the solution of the optimization problem through maximizing Eq. (3) with respect to the Lagrange multipliers. In the resulting function, there is a dot product of two vectors x_i and x_j . In practice, it is difficult to find out a linear function $f(x)$ for problems involving large dataset. Therefore, we need to map the 1-dimensional problem into a higher dimensional feature space where it is easier to find a linear function similar to $f(x)$ in a lower dimensional space. Fortunately, it is not necessary to express explicitly the

mapping during the computation. Actually, the dot product can be replaced by a kernel function $K(x_i, x_j)$. Putting Eq. (4) into Eq. (1), and taking into consideration the kernel function, the decision function can be alternated as:

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

and the optimization problem becomes: Maximize the quadratic function:

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

under the constraints (5) and (6). Normally, only a certain parts of the samples satisfy the property of $\alpha_i^* - \alpha_i \neq 0$, these parts are called support vectors (SVs).

There are four frequently used kernel functions, the linear function, the polynomial function, the radial basis function (RBF) and the sigmoid function. They represent different decision shapes in the feature space. In our work, we chose RBF in the training process because it has been tested to be proper for a couple of industrial applications [11]. RBF is also called the Gaussian kernel, it has the form

$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ where $\gamma > 0$ is the kernel parameter.

Sequential minimal optimization (SMO) method was proposed in [16] to solve Eq. (8) instead of standard quadratic problem (QP) techniques. The main idea of SMO is to divide the entire optimization problem into sub-problems and solve each sub-problem independently. In each step, two Lagrange multipliers are optimized analytically until the Karush-Kuhn-Tucher (KKT) conditions are satisfied. Which of the two Lagrange multipliers are selected in the current step is decided by a heuristics algorithm. The parameter b is re-computed in each step. The SMO method has been proved to be a better alternative to handle large datasets because it successfully solved memory load and numerical problems.

However, the problem solving is a time-consuming process when applied to large datasets. In [15], Brugger proposed a parallel approach to solve the quadratic problem by using SMO method. By profiling the performance of the training process on different datasets, the author found that the most time-consuming part is located in the kernel evaluation. Therefore, the author parallelized the kernel evaluations and gradient updates, then combined with inner sequential QP solver and distributed storage of kernel rows, and achieved linear speedup for regression. The tool Pismv was developed accordingly by the author. In our work, we applied this tool to train our model in order to accelerate the training process in high burden conditions.

In our work, we applied two performance evaluation methods. The first one is the mean squared error (MSE) which gives the average deviation of the predicted values to the measured one. Lower the MSE, better is the performance of the prediction. The second one is the squared correlation coefficient (SCC) which lies between [0,1] and gives the ratio of successfully predicted number of target values on total number of target values, i.e. how accurate the predicted values are compared to

the measured one. Higher the SCC, stronger is the evaluating ability.

3. OBTAINING HISTORICAL DATA

The main energy form in a building is the electric and gas consumption for every day use and the space heating demand in cold weather. The energy requirement for a building depends on many factors, so that it would be quite different for each building and quite different upon the time for one given building. The most important influence factors are the ambient weather conditions, building constructions and materials, inner occupants' behaviors. Therefore, as indicated in [17], it is difficult to obtain accurate consumption data from the real world, normally from measurement. Here we adopt a simulation method by using EnergyPlus [2]. We choose this software because it is a succession of the widely used building energy simulation software including BLAST and DOE-2. It is well tested and has comprehensive ability in calculating complex systems.

In this paper, an office located in Paris-Orly is simulated. The most important parameters for this building are given in Table 1. It is a simple building model with only one room and one thermal zone. The occupants' behaviors, e.g. the people using of facilities, air ventilation, are scheduled as normal office hours. We suppose that the space is heated by district heating, which means that the inlet temperature of the heating system keeps constant. We decided to do this simulation with really recorded weather data from November 1st to March 31th because this duration is a typical heating season in this location and we will analyze how much district heating energy would be consumed by this building. The weather conditions contain solar radiation, dry bulb temperature, relative humidity and so on. To indicate the weather data, the dry bulb air temperatures of the first 20 days in January and July are plotted in Figure 1.

Table 1: Information 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°
Walls	11N Stucco, 0.0253m 8IN Concrete HW, 0.2033m Insulation, 0.0679m
Fenestration surface	14m ² for each wall
Thermal Zones	1
Number of people	14
Air infiltration	0.0348 m ³ /s
Heating type	District heating
Cooling type	Heating, ventilating, and air conditioning (windowAirConditioner)
Other facilities	Light, Water heater

The output of the simulation consists of hourly recorded data related to the energy transformation of the building. These data are preprocessed in order to extract useful information and reorganized according to the format required by the analyzing tool. In our model training step, we take district heating demand as the target variable and 25 other variables as the

features including day type which indicates if the current day is holiday or not, hourly weather conditions, zone mean air temperatures, infiltration volume, heat gain through each window, heat gain through lights and people, zone internal total heat gain.

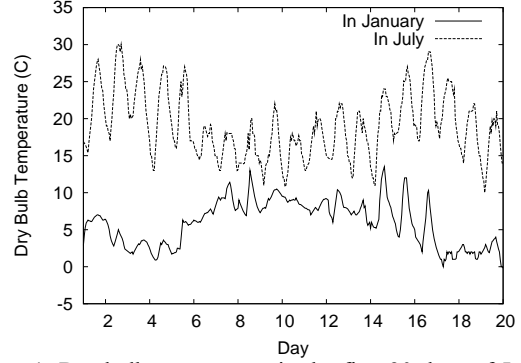


Figure 1: Dry bulb temperature in the first 20 days of January and July.

The energy consumption data for multiple buildings is generated based on the parameters of the above single one. An interface was created to automatically simulate multiple buildings under EnergyPlus. The input file is divided into two parts: an alterable part containing parameters which would be different for each building and a stable part containing parameters which would be kept the same for every building. For example, the building constructions are in the alterable part while weather data is in the stable part. For the alterable parameters, the values are obtained by a stochastic method in a reasonable range. The two parts are then combined together to form the final input file. After successfully simulating one building, we update the alterable part and repeat the simulation process for a new one. In order to analyze multiple buildings, it is necessary to put the output results of all buildings into one single output file in the proper format. The whole controlling flow is presented in Figure 2.

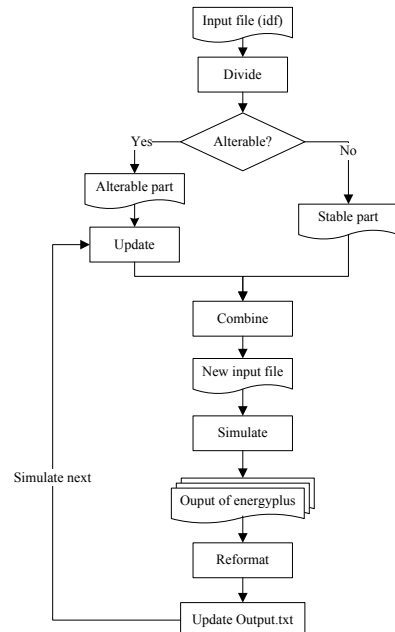


Figure 2: Flow chart of generating energy consumption for multiple buildings.

4. EXPERIMENTS AND NUMERICAL RESULTS

The dataset is divided into two parts, one is used for training models which indicates the dependence of the target on the features and the other one is used for testing the prediction performance of the obtained models. In the testing step, the predicted target regarding to the new values of features is compared with the measured one. Before training the data by SVR, one has to scale the values linearly into a small range in order to avoid numerical problems in the calculations. We scale the training subset into the range [0,1] and then apply the scaling function to scale the testing subset. The experiments are performed on a cluster composed by two homogeneous workstations with one Gigabit Ethernet network. Each computer has 8 processors (2.5GHz), 1333MHz FSB and 4G memory. The operating system used is Linux, the kernel version is 2.6.

Optimizing the parameters of SVR is important to improve the generalization ability of the models. The best choice will give the model well performance without over-fitting problem. In our work, RBF kernel is selected to train the model, so that the parameters need to be optimized are C, γ, ε . The estimation

of γ is solved by $\gamma = 1^2 / (\sum_{i,j=1}^l \|x_i - x_j\|^2)$ as indicated in [15] and [18]. The parameters C and ε are optimized by 5-fold cross validation on randomly selected 3000 samples from the training set.

One hundred buildings located in the same place in heating season are here simulated. The energy consumption data of the first 99 buildings is taken as training set, and the data for the last building is taken as testing set. Therefore, the number of samples for training is equal to 358776 and the number of samples for testing is 3264. The demand of district heating is taken as the target and the other 28 parameters such as weather conditions, the building structures, are taken as features in the analyzing process. The parameters for SVR and RBF kernel are set as $C = 4, \gamma = 0.3179$ and $\varepsilon = 0.01$. As the result shows, the number of support vectors are 27501, MSE is $5.01e-5$ and SCC is 0.997639. The predicted target compared with the measured one for the first 100 samples is plotted in Figure 3. The model shows very good generalization performance in the result. It proves that SVR with optimized parameters has a comprehensive ability in predicting energy consumption of a new building if we provide enough historical information to the model training.

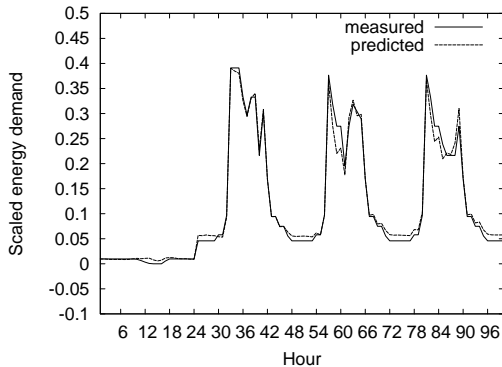


Figure 3: Measured and predicted district heating demand for the last building in heating season.

In order to improve the precision of the prediction, one has to collect more historical samples for training. However, the calculating time on processors would increase very fast while the data size is growing. It is a trade-off between the modeling time and the precision. In the above experiment, the model training process needs 31.4 hours in sequential implementation of SVMs — Libsvm [19]. It is really long for analyzing 100 buildings. To make the analyzing method more practical, it is necessary to reduce the learning time on large datasets. In next experiment, we turn to a parallel approach of SVMs to train the model and test how it performs on this problem.

We analyzed the same dataset as in the above experiment on 1, 2, 4, 6, 8 processors in Pismv. The cache size is set as 256MB for each process of MPI. Every training process is repeated three times. The average running time and speedup are calculated and shown in Figure 4. From the curves we can see that the speedups of the parallel implementation on 2 and 4 processors are quite close to linear speedup. But the performance enhancement seems not obvious when the number of processors is increased to 6 and 8. That is because of the limitation of our testing environment. On one side, SMP structure inside one node can not make distributed cache approach achieving its full effect when the number of processes is big. On the other side, gigabit network connection is very slow compared to the high performance processors. However, from the performance under 2 and 4 processors, a better speedup can be expected on servers connected by higher speed devices such as infiniband.

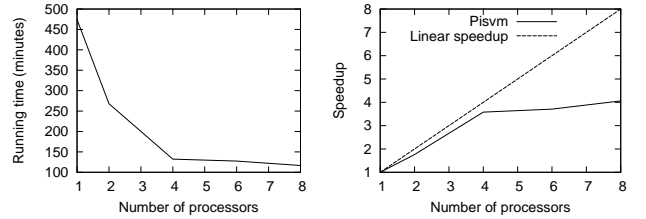


Figure 4: (a) Running time of the training process using a parallel implementation of SVMs. (b) Comparison of the speedup with a theoretical optimal linear speedup.

The results obtained in those experiments, the number of SVs, MSE and SCC, keep stable under different number of processors. The average results of parallel implementation are compared with sequential implementation in Table 2. This table indicates that, the prediction performance of the parallel solution on large datasets is quite close to the sequential one performed on smaller datasets. The parallel solution can be applied to predict the building energy performances in more complex situations which will allow longer time duration and more features and sample buildings.

Table 2: Comparison of parallel and sequential implementations: The number of Support Vectors (SVs), Mean Squared Error (MSE), Squared Correlation Coefficient (SCC).

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

5. CONCLUSION

This paper introduces a simulation approach to collect enough historical time series data for multiple buildings' energy consumption. A statistical learning method is then applied to predict the energy behavior in a completely new building. A parallel implementation of support vector regression with RBF kernel is applied to analyze large amounts of energy consumption data. Experimental results demonstrate that SVR with optimized parameters show a very good generalization ability on such problems. The good performance of the parallel implementation enhances the potential application of SVR on more complex models involving large amounts of samples or features.

REFERENCES

- [1] M.S. Al-Homoud. Computer-aided building energy analysis techniques. *Building and Environment*, 36:421–433, 2001.
- [2] EnergyPlus. Available on line at: <http://www.EnergyPlus.gov>, 2008. (Accessed July 2009).
- [3] D. Ruch and D.E. Claridge. A four parameter change-point model for predicting energy consumption in commercial buildings. *ASME Journal of Solar Energy Engineering*, 114(2):77–83, 1992.
- [4] S. Katipamula, T.A. Reddy and D.E. Claridge. Multivariate regression modeling. *ASME Journal of Solar Energy Engineering*, 120:177–184, 1998.
- [5] A. Kimbara, S. Kurosu, R. Endo, K. Kamimura, T. Atsuba and A. Yamada. On-line prediction for load profile of an air-conditioning system. *ASHRAE Transactions*, 101(2):198–207, 1995.
- [6] G. Mihalakakou, M. Santamouris and A. Tsangrassoulis. On the energy consumption in residential buildings. *Energy and Buildings*, 34:727–736, 2002.
- [7] J.F. Kreider, D.E. Clardige and P. Curtiss. Building energy use prediction and system identification using recurrent neural networks. *Journal of solar energy engineering*, 117(3):161–166, 1995.
- [8] P.A. González and J.M. Zamarreo. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37(6):595–601, 2005.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] V.N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [15] D. Brugger. Parallel support vector machines. In *VLSI-SoC: Proceedings of the IFIP International Conference on Very Large Scale Integration of System on Chip*. Springer, 10 2007.
- [16] J.C. Platt. *Fast training of support vector machines using sequential minimal optimization*. MIT Press, Cambridge, MA, USA, 1999.
- [17] L. Pérez-Lombard, J. Ortiz and C. Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.
- [18] 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.
- [19] 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).