

Machine Learning Techniques for Short-Term Electric Power Demand Prediction

Fernando Mateo¹, Juan J. Carrasco¹, Mónica Millán-Giraldo^{1 2}, Abderrahim Sellami¹, Pablo Escandell-Montero¹, José M. Martínez-Martínez¹ and Emilio Soria-Olivas¹ *

- 1- University of Valencia - Intelligent Data Analysis Laboratory
Av. Universitat S/N, 46100 Burjassot - Valencia, Spain
- 2- University Jaume I - Institute of New Imaging Technologies
Av. Sos Baynat S/N, 12071 - Castelló de la Plana, Spain

Abstract. Since several years ago, power consumption forecast has attracted considerable attention from the scientific community. Although there exist several works that deal with this issue, it remains open. The good management of energy consumption in HVAC (*Heating, Ventilation and Air Conditioning*) systems for large households and public buildings may benefit from a sustainable development in terms of economy and environmental preservation. In this paper, several Machine Learning techniques are evaluated and compared with a linear technique (Robust Multiple Linear Regression) and a naïve method. All methods have been applied to five buildings of the University of León (Spain), the results indicate non-linear techniques outperform the linear one in most scenarios.

1 Introduction

Nowadays, power load forecasting is a necessary process to achieve efficient resource management in power systems in the smart grid, and in industrial, commercial and residential buildings [1]. It has attracted much more attention with the recent deregulation of the electricity industry around the world [2]. A wide variety of models have been developed for power load forecasting, such as time series approaches [3], regression analysis [4], Artificial Neural Networks (ANN) methods [5] and, more recently, Support Vector Machines (SVM) [6].

In this work, we evaluate several Machine Learning techniques and linear models for load prediction. For this purpose, we use real data of climatic variables and power consumption measured in five buildings intended for different activities in the University of León (Spain). We are particularly interested in the next-hour prediction horizon. This paper has the following structure. In Section 2 we describe the collected data. A brief introduction of the Machine Learning methods used in this work is presented in Section 3. The experiments and results are described in Section 4. Finally, in Section 5, the main conclusions of the work are presented.

*This work was financed by the Spanish *Ministerio de Economía y Competitividad* (MINECO) and FEDER funds in the framework of the project EFIS: *Un Sistema Inteligente Adaptativo para la Gestión Eficiente de la Energía en Grandes Edificios*, with reference IPT-2011-0962-920000. The data used for the tests was provided by the University of León.

2 Data Description

Climatic variables and power consumption of a number of buildings in the University of León (Spain) have been used for the predictions. This database contains 32 buildings. Five of them with different features are chosen to evaluate the behavior of the models. Table 1 shows the list of variables used for the models evaluation (10 variables), where time and weather variables are common to all buildings. The used dataset contains samples of the variables from March 2011 to March 2012 (13 months). Each of these variables has been sampled with a frequency of 2 minutes, constituting a dataset of 280800 samples. An averaging is performed every 30 samples, leading to a sampling rate of one hour.

Variable	Description
DayM	Day of the month
Month	Month
Time	Official time
OT	Outside temperature ($^{\circ}\text{C}$)
RH	Relative humidity (%)
SR	Solar radiation (W/m^2)
WD	Working day
AP	Active Power (kW/h)
SAP	Std Active Power (kW/h)
PT	Predicted Active Power (kW/h)

Table 1: Variables used in the models training and evaluation.

3 Machine Learning Methods Used

In the recent times, machine learning methods have proved to be reliable non-linear function approximators [7]. They have managed to overcome the difficulties that complex, non-linear problems pose to simpler algorithms that do not gather domain knowledge for each particular problem.

There is a wide range of possibilities when choosing a machine learning method for a regression task. Sometimes, when the relationships among variables are not very complex, or there are significant linear components, polynomial fits like the ones obtained using Multiple Linear Regression (MLR) and its robust version (RMLR) can be good and simple options [8, 9]. When more complex relationships among variables appear, universal approximators such as Artificial Neural Networks and, in particular, the *Multilayer Perceptrons* (MLP), have demonstrated robustness and good performance in a variety of scenarios [10, 11]. However, they are often slow to train and prone to converge to local minima. Besides, overfitting is sometimes difficult to control.

A recent discovery in the field of neural networks is the *Extreme Learning Machine* (ELM) algorithm, a fast way to build and train Single-hidden Layer Feedforward Networks (SLFNs). The ELM avoids to use the *Backpropagation*

algorithm and only needs to solve a linear system of equations using least-squares [12]. Another popular model that can be used both for regression and classification is the least-squares version of the *Support Vector Machine* (LS-SVM) [13] which is derived from the original SVM used for classification and has proved very good results in the recent machine learning literature. The mentioned models can be briefly described as follows:

- *Robust Multiple Linear Regression (RMLR)*: Let $X_{n \times p}$ be the input data, y the output variable and ϵ the error vector [14], the Multiple Linear Regression (MLR) fits a linear data model as follows:

$$y_i = b_0 + \sum_{k=1}^p b_k X_{(i,k)} + \epsilon_i, \quad i = 1, \dots, n \quad (1)$$

The robust version of the regression function (RMLR) uses an iteratively reweighted least-squares algorithm, with the weights at each iteration calculated by applying the bisquare function to the residuals from the previous iteration. This algorithm gives lower weight to points that do not fit well. The results are less sensitive to *outliers* in the data as compared with ordinary least-squares regression [15].

- *Multilayer Perceptron (MLP)*: It is a type of *feedforward* neural network that can contain one or more hidden layers whose activation functions are non-linear, usually sigmoid-like functions. The Universal Approximation Theorem claims that a single-hidden layer MLP is enough to provide the input-output mapping that approximates any continuous function in a given interval [16]. The MLPs are trained using the *Backpropagation* algorithm or one of its variants. This method is powerful and computationally efficient, but it is prone to convergence to local minima and, depending on the case, it can be very slow.
- *Extreme Learning Machine (ELM)*: It is an optimized learning algorithm for SLFNs [12] that addresses the MLPs learning speed problem. The ELM algorithm has the following stages:
 1. The coefficients of the hidden layer are randomly initialized.
 2. Calculate the optimal output weights using the pseudo-inverse of the hidden layer output matrix [12].

ELM learning theory shows that the hidden nodes of generalized feedforward networks need not be tuned and that they can be randomly generated according to any continuous sampling distribution. Many types of hidden nodes including additive/RBF nodes, multiplicative nodes and non neural-like nodes, can be used as long as they are piecewise non-linear. In theory, this algorithm tends to provide a good generalization performance at extremely fast learning speed [17].

- *Least Square - Support Vector Machines (LS-SVM): Support Vector Machines*, in their classical approach for classification, try to convert a non-separable problem in the input space into a separable one in the so-called *feature* space, by finding the optimal separation hyperplane in a higher dimensional space [18]. The least-squares formulation of SVM, abbreviated as LS-SVM [13], was proposed to find the solution of the system of linear equations present in the classical SVM by means of least squares, instead of using convex quadratic programming (QP). The resulting method has similar properties to MLP and it can also be applied for regression.

4 Experiments and Results

The methods introduced in Section 3 were applied to the data described in Section 2. All input variables, Table 1, were used in the models training and evaluation. A naïve approach has been added to the table as well. Naïve forecasts are the most cost-effective and efficient objective forecasting models, and provide a benchmark against which more sophisticated models can be compared. This approach assigns the current value of the variable to predict to the next hour forecast. The error for all methods was measured in terms of Mean Absolute Error (MAE), Mean Error (ME) and Root Mean Squared Error (RMSE). The computational times correspond to a PC fitted with an Intel®Core™2 Quad CPU Q6600 @ 2.40 GHz with 4 GB RAM and running Windows 7 64-bit.

The models were trained 30 times to achieve robust results. In every training, a different random splitting of the data for training and test was carried out. The ratio of training samples was $2/3$ while the remaining data was used for testing purposes. The input data were transformed to their standard scores to produce variables with zero mean and unit variance. The standard deviation of the desired output was also calculated and averaged for the 30 trainings in order to give insight on the output fluctuation as compared to the error of the methods.

The average results are listed in Table 2 for each building. The best MAE results have been highlighted in bold. The RMLR employed the bisquare function for the calculation of the residuals [19]. The number hidden nodes for the ELMs was evaluated from 50 to 400 in steps of 50 nodes, and the best architecture was chosen. For the MLPs, a similar procedure for the selection of the number of neurons in each hidden layer, i.e. N_1 and N_2 , was followed. Combinations of layer sizes ranging from 5 to 30 neurons in steps of 5 neurons were evaluated. The best choice for the MLP size is listed in Table 2.

For the LS-SVM models, radial basis function (RBF) kernels were used. The regularization constant γ and the squared kernel bandwidth σ^2 were tuned using a 10-fold cross validation with a Coupled Simulated Annealing routine [20] to determine suitable tuning parameters and a simplex method to fine-tune them. With regard to the performance of each method, the MLPs and LS-SVMs obtained the best errors in 2 of the buildings each, while RMLR was the best method in one of the cases. LS-SVMs are all-around performers but

require a costly optimization of their hyperparameters, which is an important drawback. The naïve method lagged generally behind except for building 2 in which the simple prediction performs similarly to that of the more advanced methods. Also, RMLR has improved the results of the non-linear methods for this building, showing that this series has considerable linear content. ELMs did not perform specially well despite using a much larger number of neurons than their MLP counterparts. However, as their training times are sometimes one order of magnitude lower than those of the MLPs, this method is still an interesting option given its good trade-off between speed and accuracy.

n ^o	Measur.	RMLR	ELM	Nod.	MLP	Nod.	LS-SVM	Naïve
1	MAE	5.6278	5.5265		5.3717		4.9934	6.1562
	ME	-0.5328	0.0792		-0.0962		0.0750	-0.0375
	RMSE	8.3171	7.9342	250	7.7584	5, 15	7.3619	9.2996
	Std.	23.1022	22.9451		23.1297		23.0519	23.1213
	Time (s)	0.1160	1.0696		3.4437		4.9187	-
2	MAE	0.1900	0.2332		0.2127		0.1938	0.1905
	ME	-0.0137	-0.0029		0.0015		-0.0014	-0.0052
	RMSE	0.4863	0.4451	250	0.4451	5, 25	0.4185	0.4922
	Std.	0.9925	0.9879		0.9857		0.9942	0.9880
	Time (s)	0.2812	1.4986		4.7351		10.2600	-
3	MAE	7.7406	6.8658		4.8575		5.5433	10.0007
	ME	-1.9490	-0.0486		-0.0012		-0.0208	-0.0532
	RMSE	13.5450	9.9111	300	8.0096	5, 25	8.3904	17.3316
	Std.	55.1393	55.4307		55.4337		55.2225	55.3374
	Time (s)	0.2363	2.3411		19.0684		11.3420	-
4	MAE	3.1949	3.1772		2.7620		2.7071	3.1593
	ME	0.3872	0.0053		0.0192		-0.0005	-0.0645
	RMSE	7.0668	5.4466	300	4.9140	5, 15	4.9284	6.8073
	Std.	12.2834	12.3166		12.3194		12.1793	12.2964
	Time (s)	0.3057	1.5677		5.6266		11.4840	-
5	MAE	3.6970	3.5175		2.6614		2.9660	4.1576
	ME	-0.0649	0.0007		0.0109		-0.0274	-0.0134
	RMSE	5.4734	4.7701	300	3.6823	5, 15	4.0902	6.4496
	Std.	15.0587	15.0236		15.0230		14.9855	15.0379
	Time (s)	0.1859	1.5543		14.6037		11.6363	-

Table 2: Mean results for five buildings, sorted from 1 to 5, with the proposed methods. MAE, ME, RMSE and the standard deviation of the output are expressed in kW/h.

5 Conclusions

Power consumption forecast is very important for a sustainable development both in terms of economy and environmental preservation. This paper focuses on the use of different machine learning methods for the prediction of electric power demand in a set of large buildings, more precisely, in the University of León (Spain). The results have proved the usefulness of these methods.

The compared methods have been RMLR, ELM, MLP and LS-SVM. The results show a general advantage for non-linear techniques, such as those based in neural networks. MLPs and LS-SVM have produced the best MAE in two of the cases each, while RMLR was the best option in one of the cases. ELMs did not improve the errors of the other machine learning techniques but their computational time is much lower while enabling reasonable estimations.

References

- [1] P. R. S. Jota, V. R. B. Silva, and F. G. Jota. Building load management using cluster and statistical analyses. *International Journal of Electrical Power & Energy Systems*, 33(8):1498 – 1505, 2011.
- [2] International Energy Agency. Energy Efficiency Governance. *OECD/IEA, Paris*, 2010.
- [3] L. D. Paarmann and M. D. Najjar. Adaptive online load forecasting via time series modeling. *Electric Power Systems Research*, 32(3):219 – 225, 1995.
- [4] A. Goia, C. May, and G. Fusai. Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, 26(4):700 – 711, 2010.
- [5] Y. Wang, D. Niu, and L. Ji. Short-term power load forecasting based on IVL-BP neural network technology. *Systems Engineering Procedia*, 4(0):168 – 174, 2012.
- [6] Y. Zheng, L. Zhu, and X. Zou. Short-term load forecasting based on gaussian wavelet svm. *Energy Procedia*, 12(0):387 – 393, 2011.
- [7] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, 3rd edition, 1998.
- [8] N. Amral, C.S. Ozveren, and D. King. Short term load forecasting using multiple linear regression. In *Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International*, pages 1192 –1198, sept. 2007.
- [9] Ahmad Zia Ul-Saufie, Ahmad Shukri Yahaya, Nor Azam Ramli, and Hazrul Abdul Hamid. Robust Regression Models for Predicting PM10 Concentration in an Industrial Area. *International Journal of Engineering and Technology*, 2(3):364–370, 2012.
- [10] D. Bhattacharjee, D. Basu, M. Nasipuri, and M. Kundu. Human face recognition using fuzzy multilayer perceptron. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 14:559 – 570, 2010.
- [11] F. Djerfaf, D. Vincent, S. Robert, and A. Merzouki. Application of multilayer perceptron neural networks for predicting the permeability tensor components of thin ferrite films. *The European Physical Journal Applied Physics*, 56:30601 – 30609, 2011.
- [12] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: A new learning scheme of feedforward networks. *Neurocomputing*, 70:489–501, 2006.
- [13] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.
- [14] S. Weisberg. *Applied Linear Regression*. John Wiley, New York, second edition, 1985.
- [15] P. W. Holland and R. E. Welsch. Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics: Theory and Methods*, A6:813–827, 1977.
- [16] G. Cybenko. Approximations by superpositions of sigmoidal functions. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 1989.
- [17] G.-B. Huang, D. Wang, and Y. Lan. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2:107–122, 2011.
- [18] V. Vapnik. *Statistical learning theory*. Wiley, 1998.
- [19] J. Fox. Robust regression. Technical report, Sage, Thousand Oaks / London / New Delhi, Jan. 2002.
- [20] S. X. de Souza, J. A. K. Suykens, J. Vandewalle, and D. Bollé. Coupled simulated annealing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 40(2):320–335, 2010.