Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 6 (2020) 182–187
www.elsevier.com/locate/egyr

7th International Conference on Energy and Environment Research, ICEER 2020, 14–18 September, ISEP, Porto, Portugal

Using diverse sensors in load forecasting in an office building to support energy management

Daniel Ramos[a,b], Brigida Teixeira[a,b], Pedro Faria[a,b], Luis Gomes[a,b], Omid Abrishambaf[a,b], Zita Vale[b,*]

[a] GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Rua Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal
[b] Polytechnic of Porto, R. Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal

Received 23 October 2020; accepted 11 November 2020

Abstract

The increasing penetration of renewable energy sources led to the development of several energy management approaches. One of the main topics in this field is related to the load forecast in buildings, which can contribute to more intelligent and sustainable energy consumption. However, it is necessary to build a proper forecast model, capable of detecting an accurate consumption profile. The minimum effort to achieve this is to extract a historic with energy consumptions to use as input. Additional information should be considered in order to achieve improvements in forecasting results. This way, information regarding the day of the week is discussed as a reliable source of information that may enhance the load forecast. In this paper, two forecasting techniques, namely neural networks and support vector machine, are used to predict the energy consumption of a building for all 5 min from a period. The proposed model finds the best forecasting technique and determines if the additional information regarding the day of the week enhances the load forecast. In this case study, a period of two years and a half data with a 5-minute time interval is used. Moreover, several tests are performed for varied inputs to understand if the insights are consistent for these tests. This data has been adapted from an office building to illustrate the advantages of the proposed methodology.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Energy and Environment Research, ICEER, 2020.

Keywords: Clustering; Data mining; Fuzzy C-means; Typical load profile; Unsupervised learning

1. Introduction

The improvement of the efficient use of energy is correlated with the adequation and integration of smart grids [1]. Demand Response (DR) has an essential role in smart grids since it allows to change the power consumption to better match the demand to the supply. The interest is to study conditions that may convince consumers to reduce

* Corresponding author.
E-mail address: zav@isep.ipp.pt (Z. Vale).

https://doi.org/10.1016/j.egyr.2020.11.100

Peer-review under responsibility of the scientific committee of the 7th International Conference on Energy and Environment Research, ICEER, 2020.

their consumption on particular periods. Although this is an important question, in other programs, consumers are notified by the price signals in real-time hence they can modify the consumption and reduce electricity costs [2].

For instance, the electrical building measures real-time data with information on electricity consumptions. Monitored data is kept persistent and accessible in order to be used by forecast techniques and prevent energy waste. To this end, several forecasting techniques can be used. Artificial Neural Networks (ANN) is featured by a model composed of neurons structured in layers with links and weights values connecting them. The architecture of ANN can change; however, there is an obligation for information to start in an input layer and results are provided in an output layer and may or not have a sequential hidden layer between the input and the output used for operations. Layers send the output of the layer to the next layer, which they turn into an input. Support Vector Machine (SVM) works with a set of points placed in a dimension space. It establishes margins separating them through margins featuring several groups containing in each of these groups similarities between these [3].

These forecasting techniques were applied to a specific problem in the context of the energy domain, consistent with Short Time Load Forecasting (STLF). A forecasting application integrated with the electric load of NEPOOL region takes into account the use of an ANN architecture model [4]. A different forecasting application integrated with the prediction of wind speed and power generation is based in an SVM model [5]. The performance of the forecasting techniques ANN and SVM are researched in the context of load forecasting with the application of two models. The first one is integrated with the maximal load of forecasting day and the second one acts as an hourly load predictor [6]. The BigDEAL team proposes a methodology in NPower Forecasting Challenge, where several forecasting techniques, including ANN, are used to predict the energy usage of a group of customers day-ahead [7]. The biological systems and natural behavior is a field of study integrated into the 7th International Conference on Modern Circuits and Systems Technologies used to predict future behavior with the support of the ANN and SVM forecasting techniques [8]. The energy efficiency of the buildings is a concern of the electricity sector which focus on the power distribution network from the system operations to the end-users. Moreover, energy efficiency is improved with other technologies, including SCADA and IoT systems [9]. More precisely, these technologies are specialists in the monitor and manipulation of energy consumption data. This data is useful for forecasting techniques which provide benefit to electricity markets and policy formulations [10].

The energy forecasting methods are used as advantageous for future requirements concerning infrastructure design and planning [11]. The review presented in [12], has mentioned three approaches to building energy consumption and modeling, which are the following: physics-based, data-driven and hybrid models. While these three methods may have advantages and disadvantages of their own, the data-driven method is shown to be the most adequate solution in the context of merge building in the smart grids. Energy forecasting in the building is performed by relying on energy-related information mainly measured by smart meters. The lacking of forecasting models validation is a crucial aspect, namely in residential and office buildings [13]. Real-time automatic energy forecasts may be used for data monitoring in the context of measurement of electricity in buildings optimizing the management of energy [14].

In the present paper, the proposed methodology aims to provide a solution to improve the forecast of electricity consumption and forecasting algorithms. ANN and SVM are forecasting techniques, acknowledged as useful in decision-making problems. This paper acknowledges aspects concerning these forecasting techniques, including dealing with missing data. Moreover, adding information about the weekdays is discussed in order to decide if this information may contribute to improving the energy consumption forecasting in the building. The forecast service is run for each period of 5 min.

After this introduction, the proposed method is described in Section 2, with details about each stage. A case study is described in Section 3 to validate and test the performance of the model. The obtained results are presented in Section 4. Finally, Section 5 presents the main conclusions of the work and its future.

2. Approach

This section explains the required steps included in the proposed architecture. Several aspects are considered, such as raw data present in the database, a reduction process that transforms the raw data into information adapted for the problem at hand, the splitting of the training and test data and the forecast running integrated with the schedule of the forecast tasks. Fig. 1 shows the diagram of the proposed methodology. Using this methodology allows to contribute to the efficient management of energy consumption.

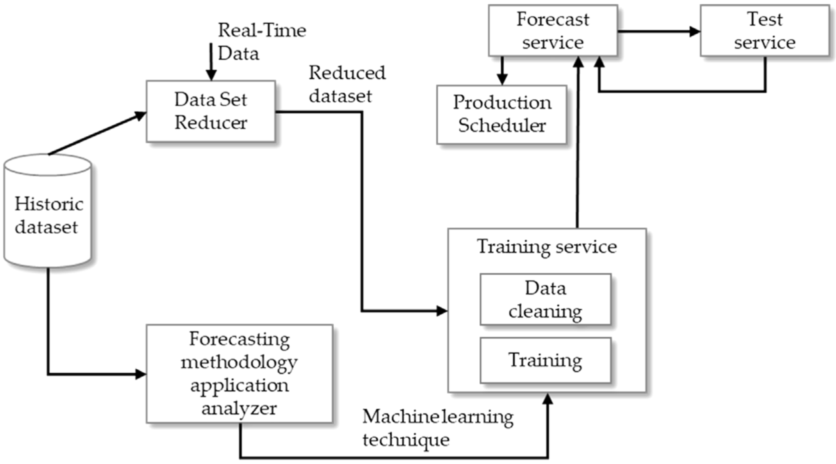

Fig. 1. Proposed two-stage methodology diagram.

Initially, the tuning process focuses on the parametrization of the involved data in the data warehouse domain evidencing two relevant aspects. The real-time data takes an active role in the first aspect, which consists of the consumptions and sensors data monitored in GECAD-ISEP associated with the N building. The first aspect is defined by a mechanism that analyzes the historical data extracted from the gathering of data and uses the extracted insights from the analysis to select the best forecasting algorithm to be used for a specific problem. The second aspect is a process for reducing data (monitored information available in the database). The process creates a sample which includes the required information to support the algorithm learning process and the targets needed for forecast, resulting in a data structure more suitable and interpretable by the forecasting algorithm. This step is crucial in the whole process since it performs accurate analysis in order to determine the parameters involved in the data warehouse domain. The obtained parameters are useful on future steps involving data forecasts tasks recalling that the accuracy of the forecast is very dependent on the data reliability and the definition of data structure. After this, the reduced version of the dataset and the selected forecast algorithm are sent to the training service. The tuning process is agnostic of which type of data it is dealing with, and it is performed only once.

The training service, in an initial phase, consists of the cleaning data process. It will perform the reorganization of the information resulting from the reduced dataset from the previous step. The new data structure has the date split in several fields involving the year, month, day of the month, day of the week, hours and minutes. This conversion helps the algorithm to perform better with time series forecast. Furthermore, the lack of information is an issue that may impact the forecast process due to the lack of observations in the training process. Therefore, missing information is added automatically following a criterion that makes sequential copies of the previous records fulling the empty fields with previously missing information. Having outliers is an issue that is described by erroneous reading made by devices that measure consumptions and sensors data. The detection of outliers takes place due to occurrences with anomaly presences in the dataset with the support of the mean and standard deviation operations. For outliers removal, the adopted strategy was replacing any outliers occurrence with the average of the records coming previously and afterwards. In the next final phase, the training reorganizes the data resulted from the cleaning process in a structure suitable for the forecast process. The training service takes action in one of the two ways: right after finishing the tuning operation, or after the system receives a new training request. The second option can be triggered only and if only the parameters were defined previously in the tuning process.

The forecast service performs predictions for all target iterations placed in the test set. The activation of the production scheduler triggers each target forecast. The forecasting algorithm used in this step is the one defined in the tuning process. This step occurs in one of the three ways: right after finishing the training service operation, after receiving a test request, or for a new iteration triggered by the production scheduler. This step finishes once the final iteration of the forecast service is triggered by the production scheduler.

The test service calculates the errors associated with the forecast target obtained in the forecast service. The errors obtained are calculated with three possible metrics: Weight Absolute Percentage Error (WAPE), Square Mean

Absolute Percentage Error (SMAPE) and Root Mean Square Percentage Error (RMSPE). WAPE is a regression metrics based on MAPE (Mean Absolute Percentage Error) being this last metrics defined by the sum of individual absolute errors divided by the demand featuring the difference between the forecasts and the actual values. WAPE indicator is provided as an alternative that has two main advantages which are overcoming high errors during a low-demand period and avoid divisions by zero. These two issues are avoided in WAPE by summing all absolute errors and demands before dividing both counterparts instead of calculating the sum of errors placed on individual periods. More precisely, this ensures that the sum of demands placed in the denominator is above zero avoiding scenarios with low and null demands which result respectively in high errors and singularity problems. Although WAPE measures the average magnitude of the errors in a set of predictions it does not consider their distance. Despite, this it overcomes the issue of high errors given by low demand values.

SMAPE is a regression metrics also used to overcome the issues presented in MAPE being calculated by using the average of the absolute value of the demand and the absolute value of the forecast in the denominator. More precisely, SMAPE is described by the absolute difference between the demand and the forecast divided by half of the sum of the absolute values provided by the demand and the forecast. Although SMAPE measures the average magnitude of the errors in a set of predictions, it does not consider their distance.

RMSPE represents the square root of the average squared error, which is described by the difference square of the difference between the actual values and the forecasts. This metrics has additional benefit compared to WAPE and SMAPE in the magnitude of the errors considering the average in a set of predictions while also considering their distance supported by the calculation of the squared error. However, this metric has the disadvantage of giving more importance to the biggest errors, determining that the biggest error placed in an individual period may be enough to determine the biggest RMSPE. The use of tree error metrics allows for understanding the obtained forecasts by analyzing their results from different perspectives.

3. Case study

In order to provide additional information needed for the forecast, the reference of the day of the week is added to the dataset and included in the process illustrated in Fig. 1. Although by adding this information allows the algorithm to build a better model, it is possible that it also increases the overfitting of data. The uncertainty between the decision to do not risk the overfitting and to provide more information requires to perform additional tests to find the best option. This case study presents an analysis of the impact of adding the day of the week information to the model, by performing forecasts for all five minutes instances of one week (weekend included). Sensors are used to acquire specific data to the algorithm/methodology, and some data as already available. Several tests were performed, considering historical data from 22 May 2017 to 17 November 2019, and executing forecast for the period between 18 to 24 November 2019. As data input, was considered the adding or the discard of the day of the week, and a set 10, 50, 100, 150, 200 or 250 entries. The entries feature the number of consumption fields that give sequence to the resulted consumption placed in the output. The greater the number of entries, the more consumption information it will provide while also risking the overfitting issue.

4. Results

This section examines the results of each of the established scenarios, where Table 1 illustrates the result error value from each of the three metrics: WAPE, SMAPE, RMSPE. Beyond the errors, these scenarios are described considering different parameters, which are: the dimension and time horizon for the historical data and the forecast; the number of entries featuring consumption values that are supposed to provide the following output; and additional information that might or not be provided for each scenario featuring the day of the week.

The table shows that results provided by the ANN algorithm are better than those provided by the SVM algorithm. For instance, the average of SMAPE metric obtained by SVM algorithm is always higher (almost 1,8%, with some individual cases above of 2,5%) than by using ANN. In this way, the provided insights will be extracted with the support of the ANN algorithm. Additionally, scenarios which include the day of the week are lightly accurate. Although, it is possible to detect that by correlating the inputs day of the week with the number of entries, the scenario with the best accuracy is the one with ten entries and with the day if the week information. The observation that this insight is taken only for a limited number of entries is understandable as the day of the week is an additional entry that motivates the overfitting along with a high number of entries. In summary, adding more entries originates a higher error. Consumption forecast curves are analyzed with more detail in Fig. 2.

Table 1. Forecast errors.

#	Entries	Add Week Day?	WAPE ANN (%)	SMAPE ANN (%)	RMSPE ANN (%)	WAPE SVM (%)	SMAPE SVM (%)	RMSPE SVM (%)
1	10	No	7,93	6,39	13,95	8,29	6,73	14,85
2	10	Yes	7,71	6,28	13,43	8,28	6,72	14,85
3	50	No	8,60	7,06	14,07	10,97	8,95	15,22
4	50	Yes	8,69	7,06	14,70	10,97	8,95	15,22
5	100	No	10,01	8,79	15,20	12,28	10,21	16,27
6	100	Yes	9,04	7,54	13,90	12,28	10,20	16,27
7	150	No	10,30	8,91	16,15	12,54	10,62	16,13
8	150	Yes	10,69	9,38	16,13	12,54	10,62	16,13
9	200	No	10,60	8,98	16,48	13,63	11,72	17,57
10	200	Yes	10,93	9,49	17,06	13,63	11,72	17,57
11	250	No	10,58	9,05	16,00	13,51	11,60	16,62
12	250	Yes	10,82	9,41	16,56	13,51	11,60	16,62
Average			9,66	8,20	15,30	11,87	9,97	16,11

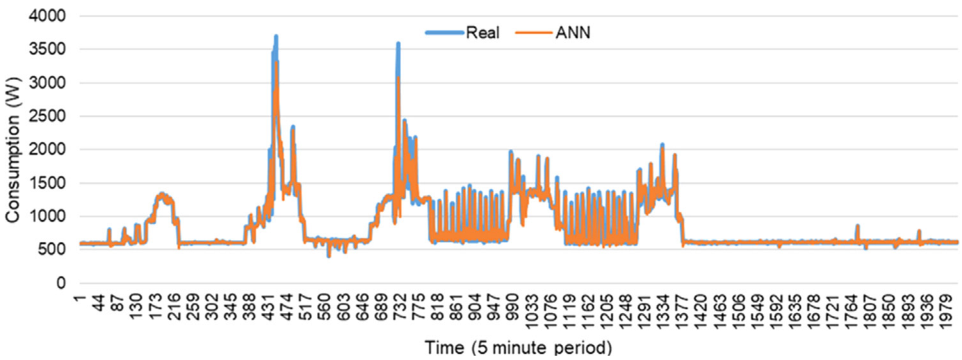

Fig. 2. Consumption forecast from 18 to 24 November 2019.

As can be seen, the forecast follows the pattern of the real consumption line. The consumption progress during the week from 18 to 24 November 2019 supports Table 1 conclusions, where the forecast observations are almost identical to the real counterparts. The used number of entries (10) allows the algorithm to better adequate to the fluctuation in the consumption, making it possible to increase the capacity of detecting peaks. The consumption starts to gain activity on Monday, reaching consumptions below 1500 W. The consumption progress placed from Tuesday to Wednesday shows that the devices measure a lot of activity reaching at a certain point a maximum above the 3500 W. Afterwards, the end of the week placed from Thursday to Friday is described by regular activity that does not exceed 2200 W. To end the consumption progress of Saturday and Sunday, this is a period of no activity at all taking into account that the useful days are from Monday to Friday. Thus, it is understandable that the measure devices monitor low activity on the weekend.

5. Conclusion

This paper details a forecasting methodology that improves the efficiency of the management of energy consumption of a building with equipped sensors that monitors the consumption data. The forecast techniques are targeted for a set of five-minute periods with the support of 2 algorithms which include ANN and SVM, with support of TensorFlow library.

The results from the case study show that the ANN algorithm has better performance, achieving more accurate forecasts. The SMAPE and RMSPE analysis demonstrate that including information with the day of the week as the input of the forecasting algorithm also provides better results. However, the number of additional entries included in input data must be limited since the studies shown a direct correlation between the number of entries and overfitting.

Thus, to achieve higher accuracy, the ANN forecast algorithm must be used, and its input must include ten entries, as well as and the information about the day of the week.

As future work, alternative options concerning the methods chosen for the forecasting process will be tested. Additionally, options concerning the exclusion of weekends and added information regarding the high and low seasons in the cleaning step should be tested.

CRediT authorship contribution statement

Daniel Ramos: Data curation, Investigation, Formal analysis, Investigation, Software, Validation, Visualization, Writing - original draft. **Brigida Teixeira:** Investigation, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Pedro Faria:** Data curation, Validation, Writing - review & editing. **Luis Gomes:** Investigation, Methodology, Validation, Writing - review & editing. **Omid Abrishambaf:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration , Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Zita Vale:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration , Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work has received funding from FEDER Funds through COMPETE program, Portugal and from National Funds through (FCT), Portugal under the projects UIDB/00760/2020, MAS-Society, Portugal (PTDC/EEI-EEE/28954/2017), CEECIND/02887/2017, and SFRH/BD/144200/2019, and from ANI, Portugal (project GREEDi).

References

[1] Faria P, Vale Z. A demand response approach to scheduling constrained load shifting. Energies 2019;12:1752, MDPI.
[2] Faria P, Vale Z. Demand response in electrical energy supply: An optimal real time pricing approach. Energy 2011;36(8):5374–84.
[3] Jozi A, Ramos D, Gomes L, Faria P, Pinto T, Vale Z. Demonstration of an energy consumption forecasting system for energy management in buildings. In: Progress in artificial intelligence. Lecture notes in computer science, vol. 11804, Cham: Springer; 2019.
[4] Singh S, Hussain S, Bazaz M. Short term load forecasting using artificial neural network. In: 2017 Fourth international conference on image information processing. 2017, p. 1–5.
[5] Zhu Z, Zhou D, Fan Z. Short term forecast of wind power generation based on SVM with pattern matching. In: 2016 IEEE international energy conference. IEEE; 2016, p. 1–6.
[6] Selakov A, et al. A comparative analysis of SVM and ANN based hybrid model for short term load forecasting. In: IEEE PES T&D 2012. 2012, p. 1–5.
[7] Xie J, Liu B, Lyu X, Hong T, Basterfield D. Combining load forecasts from independent experts. In :2015 North American power symposium. 2015, p. 1–5.
[8] Pano-Azucena A, Tlelo-Cuautle E, Tan X. Prediction of chaotic time series by using ANNs, ANFIS and SVMs. In: 7th International conference on modern circuits and systems technologies. IEEE; 2018, p. 1–4.
[9] González-Vidal A, Jiménez F, Gómez-Skarmeta AF. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. Energy Build 2019;196:71–82.
[10] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustain Cities Soc 2020;55:102052.
[11] Ahmad T, Huanxin C, Zhang D, Zhang H. Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions. Energy 2020;198:117283.
[12] Bourdeau M, Zhai X, qiang, Nefzaoui E, Guo X, Chatellier P. Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustain Cities Soc 2019;48:101533.
[13] Ha S, Tae S, Kim R. Energy demand forecast models for commercil buildings in South Korea. Energies 2019;12:2313.
[14] Gomes L, Sousa F, Vale Z. An intelligent smart plug with shared knowledge capabilities. Sensors (Basel) 2018;18(11):3961, 2018.