INTERNATIONAL BUILDING PERFORMANCE SIMULATION ASSOCIATION ENGLAND

Proceedings of BSO-V 2020
5th Building Simulation and Optimization Virtual Conference
21-22 September 2020
Loughborough, UK

Advanced machine learning techniques for predictive maintenance of HVAC subsystems based on energy consumption prediction

Desiree Arias-Requejo[1 2 3] Carlos J. Alonso-Gonzalez[4], Belarmino Pulido[4], Marcus M. Keane[1 2 3]
[1]School of Engineering, College of Science and Engineering, National University of Ireland Galway, Ireland. desiree.arias@nuigalway.ie*. marcus.keane@nuigalway.ie
[2]Informatics Research Unit for Sustainable Engineering (IRUSE) Galway, Ireland
[3]Ryan Institute, National University of Ireland Galway, Ireland
[4]Grupo de Sistemas Inteligentes, Departamento de Informática, Universidad de Valladolid, Spain. calonso@infor.uva.es, belar@infor.uva.es

Abstract

Building sector is one of the greatest energy consumers, accounting for 40% of worldwide primary energy. In the buildings, Heating, Ventilation, and Air Conditioning (HVAC) systems are major energy consumers, and their inefficient operation generates up to 50% of energy waste. Hence, techniques capable to find anomalies in these systems due to tear and wear could lead to energy and cost savings.

Nowadays the presence of smart buildings provides a valuable source of data from their meters that can be used to build data-driven models. In this work we propose to use these data in two different ways. First, we propose to use weather data to identify different working or operation modes in the buildings via hierarchical clustering. Each cluster can have different and more specific models. Second, we propose to build two kind of models for predicting the energy demand. The first one, a baseline model using historical data. The second one, based on recent data. Our guess is that estimations from both models for future energy demand will be different, being the basis for a predictive maintenance policy. We have tested our approach in a smart building in the National University of Ireland at Galway, the Alice Perry Engineering Building, with promising results.

Introduction

Between 20% and 40% of energy production is consumed by buildings worldwide (Costa et al., 2013). Buildings are the second largest emitter of CO_2, accounting for 27% of total CO_2 emissions in Europe, (International Energy Agency, 2018a). Buildings also play a major role in energy wastage. Heating, Ventilation, and Air Conditioning (HVAC) systems are major building energy consumers, and their inefficient operation generates up to 50% of energy waste (Warburton et al., 2009). System failures are responsible for most of the energy-wasting. Moreover, bigger operating and maintenance costs of modern and complex systems lead to increased operating pressures on critical building systems. Consequently, maintenance techniques should be improved by adopting new solutions that help in decision-making at the system level to optimize energy efficiency while considering maintenance costs.

Prognostics and Health Management (PHM) fits perfectly in these new maintenance strategies. PHM tracks the system current state, forecasting potential problems, or providing timely diagnoses (Katipamula et al. 2005a, 2005b). Current technology can provide many PHM functionalities like checking the system operation mode and assisting decision making in the face of abnormal situations (Satta et al., 2017). However, optimizing the system efficiency for any given operational policy under normal system behaviour is going to require the capacity to precisely estimate the short to midterm energy consumption of the system. This requirement is a must for balancing the cost of maintenance operations against the foreseen energy efficiency lose. Hence the focus of this paper is on improving energy demand forecasting and to show how those predictions can be the foundation for predictive maintenance in the context of a PHM system.

Building predictive models for energy demend forecast can be addressed mainly by two types of techniques: Physics first principles based models, and Statistical and Machine Learning models (Naug and Biswas, 2018). It is well known that Physics based models are made up essentially of mass and energy balances, combined with information about the building structure, leading to problems in their parameterization and their computational complexity, specially if they are detailed enough to be close to reality. On the data-driven side, statistical models usually rely upon some kind of distribution assumption for the data, which does not always hold. Many Machine Learning (ML) techniques do not make such an assumption, and they will be the ones used in our approach.

Since environmental conditions play a role in the energy efficiency of a building (Naug and Biswas, 2018) and may require different operation modes, we propose to associate potential different mode operations to different environmental conditions. Although in specific buildings, operation mode related variables, for instance opened/closed window shutters, may provide additional useful information, this approach has the advantage of being both simpler and of general application to any building, as long as exogenous weather-related variables measures are available.

Hence, we articulate the search for precise energy consumption prediction models in two steps using unsupervised and supervised machine learning algorithms. In this two-phase approach, we use hierarchical clustering to identify the different Air Handling Units (AHU, a specific type of HVAC system) operational states and then we apply regression models to predict the energy demand.

Different approaches have been used in ML for load forecast, being Support Vector Machines regression, Adaptive Boosting, Multi-Layer Perceptron, Recurrent Neural Networks (RNN) and, lately, Deep Learning architectures, some of the most popular. See Hrnjica B., Mehr A.D. (2020) and Amarasinghe et al. (2017) for a complete review of the methods used. We have focused on Deep Learning, and more precisely Long Short-Term Memory (LSTM) Networks for energy forecast. LSTM Networks have been specifically designed to alleviate the basic problems of RNN training: the vanishing gradient problem and the exploding gradient problem (Hochreiter & Schmidhuber, 1997). They have been quite successful in Natural Language Processing tasks, like text translation and speech recognition problems. Only recently they have been applied to short and mid-term time energy consumption prediction (Marino et al., 2016; Zhou et al., 2020) showed experimental evidence that LSTM obtained better results in both hourly and daily load predictions.

Finally, to monitor the energy efficiency of the building, we build predictive models for the global AHUs thermal energy demand, a computed Key Performance Indicator (KPI), which is directly related to the total energy demand of the AHUs.

We consider two predictive models for each operation mode. The first models are trained from far past historical data for each operation mode identified in the cluster structure. We must check that those data are from failure free operation and that maintenance operation allowed the system to achieve the desired energy efficiency. We call them baseline models. The experts of the systems must check that the predictions made with these models are acceptable for a non-fault running of the system (they have to be within an acceptable range of values). With this regard, any prediction that is lower than the baseline prediction will be acceptable and there would not be any reason to think that the system need action.

Then, for the operation mode at hand, we built a new model using current historical data. Global AHUs thermal energy demand, the computed KPI, can be estimated and compared to the expected value given the baseline model. When current energy predicted demand is significantly higher than the baseline prediction, assuming the absence of faults, a lack of efficiency due to tear and wear in different components can be presumed, and maintenance operation may be called for.

The proposed approach provides an estimation model simpler than those based on first principles models with similar or higher accuracy, accounting for the current dynamics of the plant and with a smaller computational load. The proposal can also provide an overall view of the AHU health states and can aid in their maintenance decision-making. Finally, by monitoring the system health status and the selected KPI, the energy efficiency of the system can be tracked.

Organization of the article

This paper is organized as follows: first we provide the Methodology used, including a description of the case study used to test the proposal. Results from such a test will be provided in the next Section, where we also discuss the results. Finally, we provide some Conclusions, and a glimpse of further work.

Methods

In this section first we introduce the Alice Perry building that we will use as our case study to demonstrate our proposal. The proposal will be explained later in the two following sections: first, we explain how to use hierarchical clustering to identify potential different operation or working modes in the data; second, we show our proposal of generating different data-driven models using deep learning to continuous estimation of energy demand as a first step towards predictive maintenance.

Description of the Case Study

The Alice Perry building is the home of the Engineering studies at the National University of Ireland at Galway (NUIG). It is 9 years old building, with 400 rooms (including classrooms, laboratories, and offices), a four-storeys structure, and 14,250 m^2. It was designed as a smart building integrating several green technologies for energy generation, and more than 4,000 sensors (NUIG, 2020). Our interest lies in its energy centre, which controls the HVACs, among other subsystems, by means of the Building Management System or BMS; our focus is in the main source of energy consumption: the total energy demand from the heating coils of its 11 Air Handling Units (AHU, for short).

In this complex building, with different heating and cooling technologies, there are several types of room control. Our initial guess is that the energy demand for the AHUs will depend mainly on the environmental conditions, and the usage of the different kinds of rooms, which will depend on the day of the week and the time of the day.

While the BMS will provide the energy demand for the AHUs (termed as energy demand in [kWh]), the weather station in the building will provide the “weather” dataset. We have used data from January 2018 to February 2020.

The available measurements are shown in Table 1.

In the exploration stage we eliminated several outliers with incorrect values for the barometric pressure, and rain. Additionally, we found using boxplots that air temperature, solar irradiation, and relative humidity showed more variation during spring and summer, while barometric pressure had less variation in that period. Finally, solar irradiation variation is low during winter, and wind speed is very steady during the whole year. As an example of the data variability we show in Figure 1 the evolution of the weekly mean for RH during 2018.

To cope with the complexity of this analysis our assumption is that we can identify different working or operation modes in the building, and later we can build data-driven models for each operation mode. We will describe both aspects in the following.

Table 1. Weather dataset for the Alice Perry Building.

Variable	Units	Abbreviation
Panel temperature	Celsius deg.	
Dry bulb air temperature	Celsius deg.	temp
Relative humidity	%	RH
Total solar irradiation	kW/m^2	Sir [kW/m^2] tot
Wind direction	Degrees	windirr
Avg. of 1 min. wind speed	m/s	windspeed[m/s]
Maximum 3 second gust over a 1 min of wind speed	m/s	Gust 3s avg [m/s]
Rainfall	Mm	Rain
Barometric pressure	mBar	

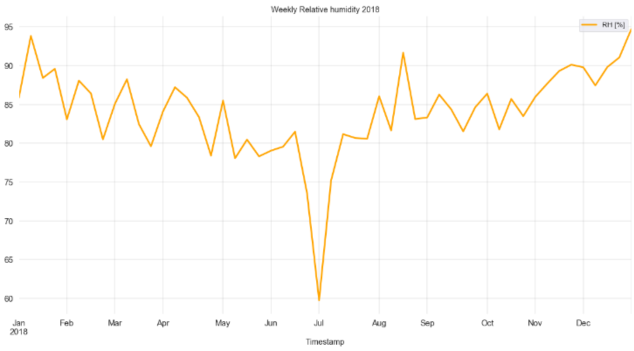

Figure 1. Weekly mean of the relative humidity in 2018.

Figure 2 illustrates the methodology used in this paper in a flow chart form. It contains the steps followed in this proposal, which will be explain in detail in the following sections.

Unsupervised learning for finding working modes

The work by Naug and Biswas (2018) showed that clustering can be used to find out different operation modes along a year attending to weather environmental variables. Arias-Requejo et al. (2020) also showed that prediction models fitted for each operation mode performs better than models trained with all the data. This paper builds on the aforementioned work (Arias-Requejo et al., 2020) and we just summarize the methodology proposed for the sake of self-containment.

In the case study at hand, the Alice Perry smart building, the weather station collects data of dry-bulb temperature, relative humidity, barometric pressure, wind speed, wind direction, total and diffuse solar irradiance, and hourly data of rainfall. From those data, five attributes where selected, namely, air temperature, relative humidity, solar irradiation, wind speed and rainfall. We used data from 2018 that were cleaned, standardized, and resampled to the weekly mean, obtaining 52 instances.

To further improve the clustering analysis, Independent Component Analysis (ICA), Hyvarinen & Oja (1997), is applied to select the relevant features for clustering. ICA is used for blind source separation and can also be used as a feature selection method. ICA proved to perform better than Principal Component Analysis, PCA, for this task. The five transformed Independent Components vectors were ranked according to their kurtosis absolute values (Ozawa and Kotani, 2000) in order to try the greatest ones as preferred candidates for the final clustering.

Hierarchical agglomerative clustering (Rokach and Maimon, 2005), was chosen to find out the relevant clusters, with the minimum variance method (Ward) -with the Euclidean metric- as similarity measure. Hierarchical clustering has the advantage of not being forced to find spherical clusters. It also helps to decide on the number of clusters because it produces a dendrogram of the grouping.

Experimental results showed that the best overall results, according to Silhouette, (Rousseeuw, 1987), Davies-Bouldin, (Davies and Bouldin, 1979), and PBM (Pakhira et al., 2004), metrics, were obtained with the three first ICA components and five clusters.

Figure 2 shows the cluster assignments obtained and used for the energy forecast models. In that graph, we can see a cluster with a single element corresponding to the last week of June (labelled as "50"). In that week the relative humidity dropped drastically, and both the solar total irradiation and the air temperature had uncommonly high values. Hence, it makes sense for that week to be isolated. This is an advantage of using Hierarchical clustering and the Ward similarity measure, that it is robust against outliers. We could interpret the cluster labelled as "1" as corresponding to winter conditions, as the first and last weeks of the year were assigned to it. The cluster "4" can be interpreted as summer because the middle part of the year was assigned to that cluster. Then, the other two clusters, labelled as "2" and "3" could be autumn and spring and different weeks during the year were assigned to them. Each of the identified clusters represent different environmental conditions in which the building could be operating under different settings and modes.

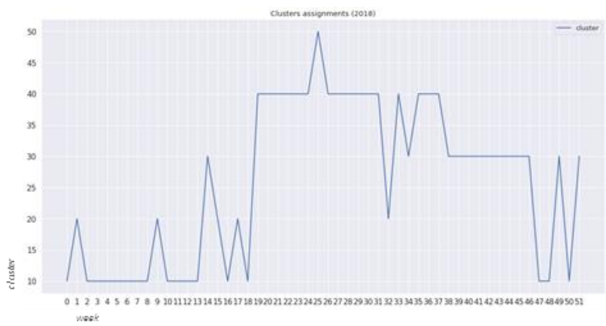

Figure 2. Weekly data cluster assignations using hierarchical clustering, 3 ICs and five clusters (k=5) for 2018 data.

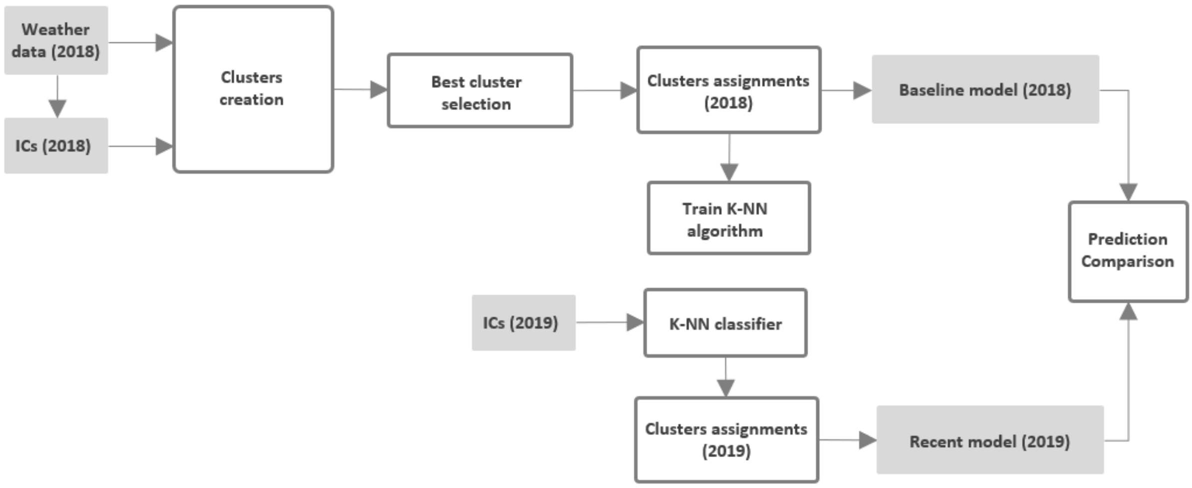

Figure 2. Flow chart of the methodology used.

In this work, we have assumed that the cluster structure remains constant along different years. Hence, we have built a classifier based on the identified cluster structure to assign the operation mode to the current data. The K-Nearest Neighbour algorithm is used for this task.

Deep Learning for building predictive models

Different approaches have been used in ML for energy load or demand forecast, mostly regression models and different configurations of Artificial Neural Networks (Feedforward, Recurrent or Deep Learning architectures). See Hrnjica B., Mehr A.D. (2020) and Amarasinghe et al. (2017) for a complete review of the methods used. Among, ANN we have focused on Deep Learning, and more precissely Long Short Term Memory (LSTM) Networks for energy demand forecast. This is specially interesting if we think about the problem of energy forecast as a time series prediction, for the short- or mid-level future, given past demand.

LSTM networks belong to the Deep Learning approach to ML. They are a special type of Recurrent Neural Networks (RNNs) with multiple hidden network layers, known as memory blocks, that containg multiple recurrently connected memory cells and three multiplicative units (input, output and forget gates), which provide write, read, and reset operations for the cells (Brownlee, 2017). LSTM networks have demonstrated to be very useful in different fields such as computer vision or speech recognition, but also for the problem known as Sequence-to-sequence prediction (Seq2seq, for short). Seq2seq prediction, given an input sequence, is able to predict an output sequence. It is a more complex problem than simply sequence prediction because it allows to have input and output sequences of varying lengths (Brownlee, 2017).

LSTMs can maintain information about their internal state for long periods of time. Their main advantage is that they do not suffer from the vanishing and exploding gradient problems as in RNNs. An additional advantage is that they can accurately model complex multivariate sequences, without pre-specifying the length of neither input nor output sequences, thus besting RNNs proposals for time series forecast. By stacking recurrent hidden neuron layers these networks are able to learn short and long term temporal correlations among temporal data series, allowing their processing at different time scales, (Malhotra et al., 2015; Marino et al., 2016).

The main goal of our work is to build models to predict the energy demand in a building for multiple time steps into the future. We propose to build a baseline model based on historical data (for instance, one year) to perform a prediction, assuming there was no fault in the system during that period of time. Additionally, we want to build a “recent” model using only early past data (for instance, during the last two or three months). If both predictions for the same period of time have statistically significant difference, we can assume that there was a significative change in the system parameters due to tear-and-wear as a part of a predictive maintenance strategy. In this work we assume that there is no other source of failure in the system (which can be assured by means of a Fault Detection and Diagnostics tool), and that we are working in the same operation mode (for instance related to similar environmental, let’s say weather, conditions).

Energy demand forecast using Standard LSTM

The goal of predicting energy demand for multiple time steps (N) in the future, given previous value for the estimated variable, and information about the current day and hour (at moment t) can be expressed as follows. We want to predict:

$$\hat{y} = \{\hat{y}_t , \dots , \hat{y}_{t+N}\} \qquad (1)$$

Given:

$$u = \{\{y_{t-1}\}, \{ d_t\}, \{h_t\}\} \qquad (2)$$

Where u is the input vector, made up of previous value for the variable to be estimated (y_{t-1}), the current day (d_t) and the current hour of the day (h_t). $\hat{y}$ is the vector output. $\hat{y}_i \in \mathbb{R}$ and $d_k \in \mathbb{N}, h_k \in \mathbb{N}$.

In our proposal, the predictions for $\hat{y}$ are generated all at once at each timestep. Hence, we do not use the $\hat{y}_{t+1}$ estimation to predict $\hat{y}_{t+2}, \ldots$

The graphical description of the stacked LSTM model as provided by Keras[1] is shown in Figure 3. The network structure is made up of the following layers:

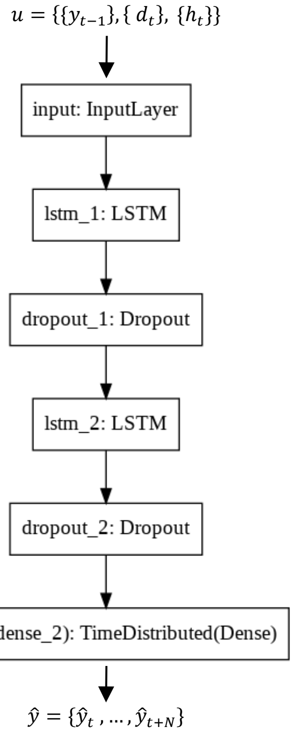

Figure 3. Stacked LSTM as provided by Keras, there is an input layer, two LSTM layers connected through a 0.5 dropout layer, and finally after the second LSTM another 0.5 dropout layer, and the time_distributed layer to generate the sequence estimation

To build the prediction models we have used a stacked LSTM architecture. We have used two LSTM layers with 50 memory cells each. We have used 60 as the timestep, 60 as the batch size, and 300 epochs for the training.

The final model provided by such structure can be summarized as:

$$\hat{y} = f(u) \qquad (3)$$

[1] Keras, https://keras.io/, is a Python platform that allows Deep Learning, acting as an interface with lower level deep learning libraries such as Tensor Flow.

Back-propagation through time (BPTT), Brownlee (2017) is used to train the models. The network is unrolled a fixed number of time steps (60 in our case). Such number is a parameter for the learning process. All the parameters in the unrolled network share the same parameters. As a consequence, once the recurrent network has 60 values, it is able to do the estimation for the next 60 time steps.

Once the network structure has been decided, the learning process consist on finding the parameters that minimize the following objective function:

$$L = \sum_{i=t+1}^{t+N} (y_i - \hat{y}_i)^2 \qquad (4)$$

Using the unrolled network standard backpropagation can be used to train the network using a gradient descend method such as Stochastic Gradient Descent (SGD). For training we use ADAM algorithm for gradient based optimization, which usually outperforms SGD, according to Marino et al. (2016). Training the baseline model took 620 seconds and training the recent model took 160 seconds.

Results on the case study

Our proposal for building both the clustering approach and the stacked LSTM network models has been done using Keras.

As it was mentioned in the Methods section, the hierarchical clustering algorithm was applied to weekly averaged weather data during 2018 for the Alice Perry Engineering Building. Results can be seen in Figure 2. We can term the cluster "10" as "winter" because most of the points fall into the winter weeks.

Later we applied the classification algorithm learnt for these data (described before) to the 2019 and early 2020 data. We obtained the assignments seen in Figure 4.

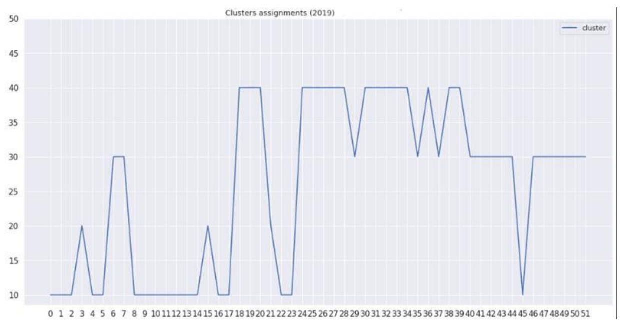

Figure 4. Weekly data cluster assignations using hierarchical clustering, 3 ICs and five clusters (k=5) for 2019.

Afterwards, data from the weeks in cluster "10" in late 2018 and early 2019 were used to build the baseline model for energy demand. Finally, we have used the weeks in late 2019 and early 2020 with label “10” to build the “recent” model.

The models were tested using data from the last three weeks in 2020 with label “10”.

The results of the prediction for both models can be seen in Figure 7.

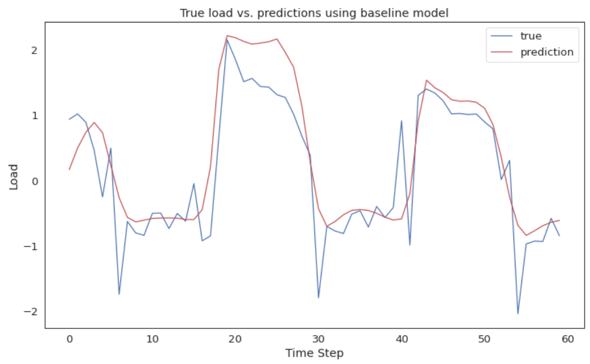

Figure 5. Energy demand forecast using the baseline model. Current demand in blue. 60 timesteps forecast in red.

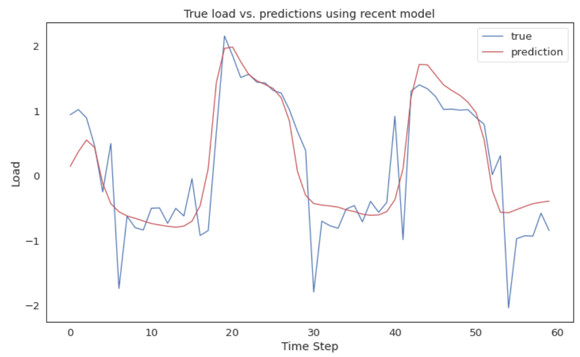

Figure 6. Energy demand forecast using the recent model. Current demand in blue. 60 timesteps forecast in red.

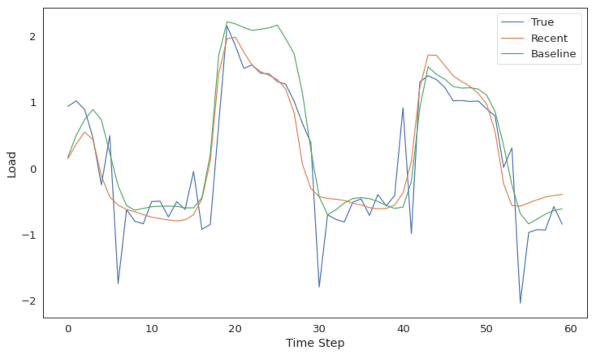

Figure 7. Difference between the true demand, and the predictions made using the baseline (historic, in green) and the recent model (in brown).

Results for these models can be summarized in Table 2. We show two metrics for the prediction. On one hand, the standard Root Mean Square Error, RMSE, value. One the other hand, we provide the Coefficient of Variation of the RMSE (CV(RMSE)) which measures the variability of the errors giving the actual values.

We can see that both the baseline and the recent model forecast for the next 60 hours are reasonably good compared to the actual energy demand. In this scenario, the baseline model predicts an accumulated energy demand for the next 60 hours greater (21,04 kwh) than the recent model (7,9 kwh), compared to the recent demand (11,78 kwh).

As said before, the baseline model has to be created with fault-free data and, for that reason, along with the system data, the experts of the system has to ensure that the data is fault-free and the demand prediction obtained with this model is acceptable. As shown in the case study presented, the recent model predicts with more accuracy the behaviour of the system and its prediction is below the baseline prediction. Considering that the baseline prediction is acceptable, the recent prediction should be acceptable too and there would not be any reason to think that the system will need any action. On the other hand, if the recent prediction had been above the baseline, it could indicate an efficiency loss in the system.

In the case study presented, we can conclude that the system does not require now maintenance action, as it would be the case if the recent estimation would be higher than the baseline prediction.

Table 2. RMSE and CV(RMSE) for baseline and recent models.

Model	RMSE	CV(RMSE)
Baseline	0.556	27.18%
Recent	0.525	25.63%

Discussion

It is rather difficult to compare our proposal against others for two reasons: first, there is no paper using deep learning techniques for the Alice Perry Building case study; second, we have not found another work combining hierarchical clustering and LSTM networks for predicting future energy demand, as a first step towards predictive maintenance for HVAC systems. For that reason we discuss several works that deals with either time series prediction using deep learning, or using LSTM for anomaly detection in time series, or using unsupervised learning for weather data to identify different operation modes in buildings.

Regarding the use of clustering to identify operation modes in a building, the proposal by Naug and Biswas (2018) is focused on optimizing control signals for energy consumption reduction, instead of predictive maintenance. Our proposal is similar because we use weather data and hierarchical clustering, but we have selected to use ICA instead of normalized values of their measurments (their clusters use air temperature, relative humidity and solar insolence) in 10 min. intervals. The way of computing if a new point belongs to a cluster is also different, because they compute the center of each cluster. We use a classification model.

Regarding the use of machine learning for predictive maintenance in HVAC buildings, our work poses a rather different approach to that in Yang et al. (2018). Their work propose to use machine learning for HVAC prognostics using data from work orders from 44 buildings. The model built is for Fault Mode and Effect Analysis. Hence, there is no energy or weather information, and the type of model is very different from ours.

Satta et al. (2017) also looks for dissimilarity between performances as a first step towards predictive maintenance for HVAC systems. However, their approach relies upon a *cohort* (set) of devices working together (16 HVAC units working together in a hospital). The need for maintenance comes from dissimilaty between the data. Their proposal is rather different from ours because they use a classification model with decision trees as base classifiers combined through AdaBoost.

Regarding the usage of LSTM architectures for energy or load prediction, main difference from other works (Maholtra et al., 2015), and (Marino et al., 2016) comes from using the clusters to figure out different operation modes.

Maholtra et al. (2015) have reported that stacked LSTM can be used as a predictive model for anomaly detection in time series, testing the proposal in different scenarios including power consumption. A stacked LSTM architecture is trained on non-anomlous data, and later is used to predict over a time interval in the future. The errors (anomalies) are modelled as gaussian distribution, providing in each time step a prediction in the future, and accumulating predictions for each time step in the future. Hence, values outside the normal distribution are considered as anomalies. Main difference with our stacked LSTM architecture is that we use not only energy data, but also the temporal information for anomaly detection. Also our detection method is different because we use predictions from the baseline and the recent model.

Marino et al. (2016) used two different LSTM-based approaches for building level load forecast: standard LSTM architecture, and LSTM-based Seq2seq architecture. The Seq2seq architecture outperformed the standard architecture for a “Individual household electric power consumption” dataset (see references therein). Those authors have later tested the use of Convolutional Neural Networks, CNN, for building-level energy load forecasting (Amarasinghe et al. 2017) providing comparable, but not much better, results. Our proposal of stacked LSTM networks is similar to their classical LSTM proposal, but the results in our proposal are better than their data. Their encoder-decoder proposal performs reasonably better than the original LSTM. Our results seems closer to that Seq2seq architecture, but wit a simpler structure. Additionally, it is rather difficult to compare both results numerically because they compute the RSME for unnormalized data, but looking at the graphics our proposal works reasonably well for 60 time steps in the future. Additionally our proposal of combining clustering to select more precise models is another difference with their proposal, although it would be rather interesting to explore the usage of CNN or hybrid approaches in the near future.

Conclusion

In this work we have demonstrated that unsupervised and supervised machine learning algorithms can be used for the purpose of energy demand prediction as a first step towards predictive management. Results obtained from our case study are not conclusive, but they are promising.

The hierarchical clustering applied to weather data, transformed using ICA, provide a clear separation in four clusters that can be thought as the four different seasons during the year, plus one cluster that clearly isolates weeks with unusual values.

The stacked LSTM architecture provides a reasonable estimation for short term energy demand in the building. Further work needs to be done to improve the estimations, for instance implementing an Encoder-Decoder architecture. It seems clear that these techniques work better with more data available, hence with another year we can improve these results.

Finally, we have proposed to build two regression models, baseline and recent, with past and recent data. We would expect that a big difference in the estimations between both models could lead to maintenance actions. In our case the estimations for the recent model are closer to the actual demand, hence it is not an indication for repair.

Acknowledgements

The data used in the case study of this paper was provided by the IRUSE team and the College of Informatics and Engineering of NUI Galway. This research work was partially funded by the NUI Galway CoEI Postgraduate Scholarship 2018, the European Commission under grant agreement number 820805 through SPHERE project, and the Intelligent Systems Group (GSI) from the Universidad de Valladolid.

References

Amarasinghe, K., Marino, D.L., and Manic, M. (2017) "Deep neural networks for energy load forecasting," 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, pp. 1483-1488, doi: 10.1109/ISIE.2017.8001465.

Arias-Requejo, D., Alonso-González, C.J., Pulido, B., Keane, M. (2020) Energy demand prediction in smart buildings using advanced machine learning techniques. eSim. Vancouver.

Brownlee, J. (2017) Long Short-Term Memory Networks with Python. Machine Learning Mastery.

Costa, A., Keane, M. M., Torrens, J. I., & Corry, E. (2013). Building operation and energy performance: Monitoring, analysis and optimisation toolkit. *Applied*

Energy, *101*, 310–316. https://doi.org/10.1016/j.apenergy.2011.10.037

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Hrnjica B., Mehr A.D. (2020) Energy Demand Forecasting Using Deep Learning. In: Al-Turjman F. (eds) Smart Cities Performability, Cognition, & Security. EAI/Springer Innovations in Communication and Computing. Springer, Cham

Hochreiter, S. and Schmidhuber, J. Long Short-term Memory. In: Neural com-putation 9 (Dec. 1997), pp. 1735{80. doi: 10.1162/neco.1997.9.8.1735.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. Neural computation, 9(7), 1483-1492.

International Energy Agency. (2018a). *CO2 Emissions from Fuel Combustion 2018 Highlights*. 166.

Katipamula, S. & Brambley, M.R. (2005) Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I, HVAC&R Research, 11:1, 3-25

Katipamula, S. & Brambley, M.R. (2005) Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part II, HVAC&R Research, 11:2, 169-187

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long Short Term Memory networks for anomaly detection in time series. In Proceedings (Vol. 89). Presses universitaires de Louvain.

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. In IECON 2016- 42nd Annual Conference of the IEEE Industrial Electronics Society (pp. 7046-7051).

Naug, A., & Biswas, G. (2018, June). Data driven methods for energy reduction in large buildings. In 2018 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 131-138. IEEE. https://doi.org/10.1109/SMARTCOMP.2018.00083

Ozawa, S., & Kotani, M. (2000). A Study of Feature Extraction and Selection Using Independent Component Analysis. 6.

Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. Pattern Recognition, 37(3), 487–501. https://doi.org/10.1016/j.patcog.2003.06.005

Rokach, L., & Maimon, O. (2005). Clustering Methods. In O. Maimon & L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook (pp. 321–352).

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Satta, R., Cavallari, S., Pomponi, E., Grasselli, D., Picheo, D., & Annis, C. (2017). A dissimilarity-based approach to predictive maintenance with application to HVAC systems. arXiv preprint arXiv:1701.03633.

Warburton, P., Butcher, K. J., & Chartered Institution of Building Services Engineers (Eds.). (2009). *Building control systems* (2. ed). CIBSE.

Yang, C., Shen, W., Chen, Q., & Gunay, B. (2018). A practical solution for HVAC prognostics: Failure mode and effects analysis in building maintenance. Journal of Building Engineering, 15, 26-32.

Zhou, C., Fang, Z., Xu, X., Zhang, X., Ding, Y., Jiang, X., & ji, Y. (2020). Using long short-term memory networks to predict energy consumption of air-conditioning systems. *Sustainable Cities and Society*, *55*, 102000. https://doi.org/10.1016/j.scs.2019.102000