Building heat disaggregation with deep learning

Kaarlo Forsman
kaarlo.forsman@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal
October 2023

Abstract

Energy retrofit projects for existing buildings aim to decrease the overall heat energy use in buildings efficiently. Informed investment decisions require disaggregated heat consumption data. However, often only the overall heat consumption data is available.

This work focused on developing a data-driven model for building heat consumption data disaggregation. A synthetic dataset was constructed from a portfolio of completed projects managed by SWECO, a Scandinavian engineering and architectural consultancy. After assessing the performance of a variety of neural network architectures, a feedforward model was found to have the best validation performance with MAE of 0.0814, RMSE of 0.1388 and r2-score of 0.704. The model was assessed to be suitable for deployment, but only in the intended preliminary assessment use case.

Keywords: machine learning, deep learning, energy, building, disaggregation

1 Introduction

Energy consumption in buildings represent a large portion of overall energy consumption and the associated greenhouse gases. In Finland, the energy consumption of buildings represents approximately 40 % of final energy consumption, with especially heating playing a significant role [1], [2]. Accurate building energy consumption forecasting, and consumption profiles are a necessary requirement for buildings energy efficiency estimation and are an integral part of energy efficiency investment decision making. Simulation models may provide accurate estimates but require a vast amounts of detailed input data about the buildings structure and the HVAC systems [3], [4]. However, in cases where detailed information about the building is not available or the nature of the design assignment does not require an extensive analysis of the building's energy balance, constructing a building simulation model might be impractical.

One of identified key tasks often faced by engineering consultancies in the building energy domain is to develop energy retrofit projects for existing buildings. Accuracy of the information about the building varies and a detailed consumption profile is rarely available. The overall heat and electricity consumption is often known, but a disaggregated consumption profile with different end-use cases differentiated is not available. Constructing a simulation model for an existing building in order to obtain approximations of disaggregated heat consumption data is time consuming and thus not ideal. Inaccurate information and resource intensive simulation model development motivate the development of a data-driven model to automate the task. [5].

In the domain of data-driven models, machine learning models are a powerful subset of algorithms that are used to leverage existing data [3]. As a part of SWECO´s project portfolio, an extensive collection of simulation models for delivered projects is available that can be utilized for data-driven model development [5].

The objective of this study is to develop a data-driven model for building heat disaggregation in the context of the SWECO project. A comprehensive review of building heating systems and machine learning techniques is conducted. Additionally, a literature review to explore data-driven models used in energy prediction, forecasting, and heat disaggregation tasks is conducted. A synthetic dataset based on IDA ICE building energy simulations is constructed, analyzed, and engineered to be utilized as the training data for data-driven model development. Data-driven models are trained and evaluated with respect to the end-use case.

2 Foundations

2.1 Heating in buildings

Energy use in buildings is a sum of complex interplay between the physical properties of the structure. Sustaining suitable conditions inside the building is done with the buildings heating, ventilation, and air conditioning systems as well as water supply system. The overall final energy use in buildings consist of electrical energy and heat energy. The scope of this work focuses on the quantities and classification of supplied heat [6].

Generally, the heat demand of the entire HVAC systems can be divided into four categories: domestic water heating, domestic hot water circulation losses, space heating and ventilation system intake air heating. The output classes of the data-driven models developed in this work are representations of these heat consumption categories.

The domestic hot water (DHW) system of a building delivers both cold and hot water for the users. DHW energy demand refers to the amount of energy that is needed to increase the temperature of the cold water to a suitable level for the buildings end-users. The hot water is distributed with a closed-loop circulation system, that allows the heated water to circulate inside the buildings water distribution system. The heat that’s transferred and lost from the hot water during the circulation process is referred to as DHV circulation losses.

Space heating & cooling systems are used to control temperature in the buildings. In this work, the focus is on heat energy use and thus the space temperature control is assumed to only be capable of adding energy into the system. Space heating aims to supply the system with equivalent amount of energy that the overall system losses are.

Ventilation systems ensure the quality of indoor air. Mechanical ventilation is often utilized for larger buildings and offers higher control over ventilation and is often utilized in larger buildings due to the higher ventilation requirements. Heating of ventilation intake air is an efficient way to reduce energy consumption. Ventilation system intake air heating refers to the energy utilized for this process [7].

2.2 Machine Learning

Data-driven models, alternatively machine learning (ML) models, consist of a family of algorithms that aim to learn the patterns from the given data. ML algorithms are often categorized into traditional algorithms and deep learning algorithms. Deep learning algorithms, also referred to as neural networks, are a family of ML algorithms that currently represent state-of-the-art performance in various benchmark tests. [8]

ML algorithm is an algorithm capable of acquiring knowledge from data. ML models can be trained to carry out a specific task and regression and classification are the most prevalent tasks in energy forecasting and prediction as well as energy disaggregation applications. The SWECO heat disaggregation problem can be classified as a regression task [9].

To establish how well a model is executing the given task, a performance metric needs to be specified. The performance measures are evaluated using a distinct set of examples, that have not been used in the training of the model to address the performance of the model in general use. Furthermore, the model performance can be validated with evaluating the performance of the model with a previously unseen set of data, often referred to as validation set [9].

Mean squared error/mean absolute error (MAE), root mean-squared error (RMSE) and normalized mean biased error are generally used metrics for regression tasks. The errors are defined formally below:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{m} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{m}} \tag{7}$$

$$NMBE = \frac{1}{\bar{y}} \times \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n} \times 100\ (\%) \tag{8}$$

Model goodness of fit can be estimated with specific performance metrics that focus on the overall model fit over a dataset. R2-score is often used and is formally defined below:

$$r2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{9}$$

2.3 Energy consumption models – review

Data-driven models have emerged as a powerful alternate to traditional physics-based models in the last decades. The popularity of data-driven models is due to their ease of development, high accuracy, and suitability for a wide array of applications. This has been further fueled by the growing interest in machine learning and deep learning models, which have been successfully applied to energy forecasting tasks. [10].

Energy disaggregation tasks involve deconstructing an aggregated data, such as overall consumption, to the original components it is originally derived from. Machine learning models have achieved good results in energy disaggregation tasks. The performance of different machine learning algorithm implementations was found to depend on the accuracy and size of available training data, the complexity of the disaggregation task and the availability of explanatory features. Neural network based algorithms were found to increase performance compared to traditional machine learning algorithms in most cases. Recurrent neural networks, a branch of neural network algorithms that utilize feedback loops, were found to further increase performance in cases, where temporal correlations were evident. Transformer based neural network architectures have presented promising results and have gained popularity in recent years [11]

Neural networks have been utilized extensively in energy forecasting & prediction tasks ML algorithms exhibited good performance in energy forecasting & prediction tasks, with neural networks achieving excellent results. Best performing architectures generally employed a recurrent unit that enables the model to extract trend patterns in previously predicted and realized consumptions [12]–[24].

3 SWECO building heat disaggregation problem and data

3.1 Synthetic dataset properties & pre-processing

The dataset was constructed from previous SWECO projects and the respective simulation models and energy calculation results. The dataset was constructed from 16 buildings that are either apartment buildings, offices or schools. Each building has at least one year of simulated consumption at the hourly scale. In total, 24 distinctive one year time-series were included in the dataset. The dataset was constructed by extracting structural constants and parameters from the simulation model and combining them with the simulation results, thus obtaining a fully synthetic dataset based on the existing SWECO project portfolio.

From each building, 22 features were gathered. Two of them, the overall heating consumption as well as the temperature are obtained from the simulation results. Rest of the features were either structural constants or use profile parameters. Additionally, several features were aggregated based on the original features, bringing the total feature count to 28. The dataset comprises of 210240 data entries in total. No outliers, missing values or otherwise problematic datapoints were found in the entire dataset. Temporal connections were found in the overall consumption time-series, with high positive autocorrelation at 24 time-steps.

3.2 Proposed model and results

The SWECO building heat disaggregation problem was formulated as a multinomial supervised regression task and the proposed models for further examination for the problem are based on the findings in the literary review.

Traditional machine learning algorithms were implemented as a first stage of statistical model development for the problem to serve as a baseline performance indicators. Multivariate linear regression as well as 2-degree and 3-degree polynomial regression models were implemented. The dataset was split for a training and test set with a ratio of 0.8. The model outputs shares of overall consumption for the four specific categories. The units for outputs and the respective error metrics are not in energy quantity or power units.

For the preliminary baseline tradition machine learning algorithm implementation, only test metrics were calculated. The 3-degree polynomial regression performed best in all the metrics used achieving an MAE of 0.078, RMSE of 0.109 & r2 -score of 0.82. Linear regression performed poorly, achieving nearly 0.15 MAE, RMSE of 0.19 and r2 -score of 0.46.

3.2.1 Neural network model implementation and accuracy

Various neural network models were implemented on the problem utilizing the full training dataset. Different network architectures were trained and evaluated in order to obtain the optimal model architecture. As well as architectural design considerations, hyperparameters values were optimized by trial-and-error model training and evaluation. Feedforward, simple RNN, GRU & LSTM architectures models were implemented.

Due to the high autocorrelation indicating a presence of temporal patterns in the consumption time-series, a series of lagged values with shifted overall consumption values were feature engineered into the training dataset. The scope of the lagged values was treated as a tunable hyperparameter value to achieve optimal results. Models were trained with both the original dataset as well as the dataset including the lagged values.

MAE, RMSE, R2-score as well as category specific MAE for the output categories were calculated as part of the evaluation of the models. The generalization performance of the models was evaluated by utilizing a distinctive validation dataset. The validation dataset is constructed of three standardized buildings that are used in SWECO internal tool performance evaluations.

Tanh, sigmoid and ReLU activation functions were tested and ReLU was found to achieve the best performance. Sigmoid, softmax & linear activations were tested for the output layer and sigmoid was found to achieve best performance. The train-test split was initiated at 0.8 and was increased to 0.95 for final hyperparameter tuning. Adaptive Adam optimizer was used.

Data was scaled to increase model performance and different scaling algorithms were tested. Power transformer scaling was found to allow models to achieve best performance. Various dropout probabilities were tested for various network architectures to survey the optimal dropout probability. A probability of 0.4 produced best results in terms of validation performance.

Table 1. Validation performance of best performing RNN, feedforward network with lagged values and feedforward network without lagged values.

Performance metric	**LSTM**	**FFN Lagged values**	**FFN**
Model description	*200-unit LSTM, 4 x 200-unit feedforward layers*	*4 x 200-unit feedforward layers. Lagged values for overall consumption.*	*4 x 200-unit feedforward layers. No aggregated lagged values.*
MAE	0.191	0.0858	0.0814
RMSE	0.265	0.1482	0.1388
r2-score	-	0.65	0.704
Category specific MAE			
Space heating	36.0 %	14.55 %	12.49 %
Ventilation intake air heating	36.1 %	11.33 %	10.54 %
Domestic hot water heating	36.1 %	5.79 %	6.49 %
Domestic hot water circulation losses	36.4 %	2.65 %	3.05 %

Recurrent neural network architectures failed to achieve comparable performance when compared to the feedforward networks with and without lagged value vision. The poor performance of recurrent neural networks despite the autoregressive nature of the data can be attributed to the insufficient amount of data compared to the significantly large complexity of the recurrent neural architectures [25].

Best performing feedforward network architecture with dataset not containing the lagged values was a four-layer deep network with 200 neurons in each layer. Dropout layer with a dropout probability of 0.4 was implemented after each feedforward layer. Batch size of 64 was used, MAE was used as the loss algorithm and the model was trained for 100 epochs. Data was scaled with power transformation Yeo-Johnson transformation. The model achieved a validation MAE of 0.08645, RMSE of 0.1472 and r2-score of 0.67 with category specific mean absolute errors of

13.9 % for space heating, 11.64 % for ventilation and intake air heating, 6.35 % for domestic hot water heating and 3.16 % for domestic hot water circulation losses.

The hyperparameter and architecture search was done individually for both lagged dataset and the non-lagged dataset. In addition to the hyperparameters described in the case of the non-lagged dataset, the number of shifts included in the lagged values was 30. The model achieved a validation MAE of 0.08611, RMSE of 0.1485 and r2-score of 0.66 with category specific mean absolute errors of 14.61 % for space heating, 11.01 % for ventilation and intake air heating, 5.77 % for domestic hot water heating and 3.05 % for domestic hot water circulation losses.

After determining the best performing model and associated hyperparameter values, the final model was trained on 0 % test set allocation leveraging the entire training set for model training. The feedforward model with and without the lagged obtained best results. The feedforward model trained on the entire dataset without the added lagged values achieved a validation MAE of 0.0814, RMSE of 0.1388 and r2-score of 0.704. Category specific mean absolute validation errors for the model were 12.49 % for space heating, 10.54 % for ventilation intake air heating, 6.49 % for domestic hot water heating and 3.05 % for domestic hot water circulation losses. The NMBE was in the magnitude of $\ll n \times 10^{-10}$.

3.2.2 Model performance with respect to end-use case & industry standards

The SWECO building heat disaggregation model is intended as a preliminary assessment tool for the evaluation of previously unseen buildings heat use. Resulting model is not intended to be used as the most comprehensive energy calculation tool since the more interpretable and accurate simulation software is more suitable for such tasks.

The overall validation performance of the model is sufficient for the defined end-use. MAE of 0.0814, RMSE of 0.1388 and r2-score of 0.704 indicate that the model is suitable for preliminary assessment use. Of the output categories, space heating and ventilation intake air heating have the highest mean absolute error. The validation performance metrics of the final model can be found in Table 2

Table 2. Final model validation performance.

Performance metric	Model validation performance
MAE	0.0814
RMSE	0.1388
r2-score	0.704
Category specific MAE	
Space heating	12.49 %
Ventilation intake air heating	10.54 %
Domestic hot water heating	6.49 %
Domestic hot water circulation losses	3.05 %

FEMP criteria, IPMVP and ASHRAE guideline 14 provide methodologies for assessment of energy prediction models. The final model performance meets the criteria, apart from the r2 -score recommendations by ASHRAE and IPMVP.

The prediction errors of the model are of similar or smaller magnitude than the expected quality of data in real-life cases, with deviation accuracies of 10 % expected to be common [5]. However, the inherit inaccuracies in the synthetic dataset and high maximum errors should be considered if the model is deployed in other applications.

4 Conclusions

The aim of this work was to develop a data-driven model for a building heat disaggregation task. The foundations of building heating systems and machine learning were reviewed. Literary review of data-driven models examined in the energy prediction and forecasting as well as heat disaggregation tasks was conducted. The findings were assessed with respect to the SWECO building heat disaggregation problem and key design factors for the model development were identified. A dataset was constructed from IDA ICE simulations from the SWECO project archive. The dataset was analyzed and engineered to serve as a training dataset for data-driven models. Traditional and deep learning algorithms were trained and evaluated, and the final model performance was assessed with respect to the end-use case.

The literary review showed that a variety of machine learning algorithms have been deployed successfully in the energy domain. Several promising data-driven algorithms were identified, and neural networks showed promising results in both energy disaggregation as well as energy prediction and forecasting tasks. Recurrent neural networks with feedback connections were found to outperform feedforward networks in most cases with especially LSTM-networks have showing promising results.

The training data of the model was constructed from building simulations sourced from SWECO project archives. Autoregression analysis confirmed a presence of strong temporal correlations for the energy consumption, which is in line with energy time-series data in general. Dataset was aggregated to contain lagged values of the overall consumption to allow models without feedback connections to leverage the temporal correlations. Traditional machine learning algorithms were trained and evaluated as baseline models to assess the suitability of data-driven models for the task.

A variety of neural network architectures were trained and evaluated. Network architectures and hyperparameter values were tried and tested to obtain best performing model structure. Recurrent neural networks did not achieve a performance comparable to less-complex feedforward neural networks.

The final model had four hidden layers and an output layer, with 200 neurons in each hidden layer. Dropout with a dropout probability of 0.4 was included after each hidden layer. ReLU activation was used for the hidden layers and softmax for the output layer. Model achieved a validation MAE of 0.0814, RMSE of 0.1388 and r2-score of 0.704. Space heating and ventilation intake air heating had highest output category specific MAE's. The maximum errors made by the model over the entire validation dataset were high. The high maximum errors and relatively low r2-score need to be further considered if alternative end-use cases are considered for the model.

The model was assessed to perform the SWECO heat disaggregation task sufficiently with respect to the preliminary assessment nature of the end-use case of the model. Expanding the relatively small dataset is identified as key factor in increasing data-driven model performance in this and other related tasks. The inclusion of real-life data into the currently entirely synthetic dataset is also identified as a key factor for increasing model generalization performance. Computational costs for models trained in the work are stated to be marginal and re-evaluating model architectures and hyperparameters values is recommended if the dataset is expanded

References

[1] Motiva, “Rakentaminen ja rakennukset.”

[2] Statistics Finland, “Statistics Finland’s statistical database - Energy,” 2022.

[3] J. Runge and R. Zmeureanu, “Forecasting energy use in buildings using artificial neural networks: A review,” *Energies*, vol. 12, no. 17. MDPI AG, Aug. 23, 2019. doi: 10.3390/en12173254.

[4] A. Vadiee and L. Gustavsson, “A comparison between four dynamic energy modeling tools for simulation of space heating demand of buildings.”

[5] SWECO, “Joni Hilpinen, kehityspäällikkö,” *Discussions about energy modelling in engineering consulting.* 2023.

[6] J. D. Racusin, *Essential building science : understanding energy and moisture in high performance house design*. in Sustainable Building Essentials. Gabriola, BC: New Society Publishers, 2016.

[7] Ympäristöministeriön asetus, “Rakennuksen energiankulutuksen ja lämmitystehontarpeen laskenta,” Helsinki, 2007.

[8] B. Macukow, “Neural networks-state of art, brief history, basic models and architecture,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2016, pp. 3–14. doi: 10.1007/978-3-319-45378-1_1.

[9] Goodfellow Ian, Bengio Yoshua, and Courville Aaron, *Deep Learning*. Massachutes: MIT Press, 2016.

[10] Z. Wang, T. Hong, and M. A. Piette, “Building thermal load prediction through shallow machine learning and deep learning,” *Appl Energy*, vol. 263, Apr. 2020, doi: 10.1016/j.apenergy.2020.114683.

[11] N. Zaeri, A. Ashouri, H. B. Gunay, and T. Abuimara, “Disaggregation of electricity and heating consumption in commercial buildings with building automation system data,” *Energy Build*, vol. 258, Mar. 2022, doi: 10.1016/j.enbuild.2021.111791.

[12] H. Y. Park, B. H. Lee, J. H. Son, and H. S. Ahn, “A comparison of neural network-based methods for load forecasting with selected input candidates,” in *Proceedings of the IEEE International Conference on Industrial Technology*, Institute of Electrical and Electronics Engineers Inc., Apr. 2017, pp. 1100–1105. doi: 10.1109/ICIT.2017.7915516.

[13] M. Mordjaoui, S. Haddad, A. Medoued, and A. Laouafi, “Electric load forecasting by using dynamic neural network,” *Int J Hydrogen Energy*, vol. 42, no. 28, pp. 17655–17663, Jul. 2017, doi: 10.1016/j.ijhydene.2017.03.101.

[14] J. Moon, S. Park, S. Rho, and E. Hwang, “A comparative analysis of artificial neural network architectures for building energy consumption forecasting,” *Int J Distrib Sens Netw*, vol. 15, no. 9, Sep. 2019, doi: 10.1177/1550147719877616.

[15] C. Fan, F. Xiao, and Y. Zhao, “A short-term building cooling load prediction method using deep learning algorithms,” *Appl Energy*, vol. 195, pp. 222–233, 2017, doi: 10.1016/j.apenergy.2017.03.064.

[16] Ö. F. Ertugrul, “Forecasting electricity load by a novel recurrent extreme learning machines approach,” *International Journal of Electrical Power and Energy Systems*, vol. 78, pp. 429–435, Jun. 2016, doi: 10.1016/j.ijepes.2015.12.006.

[17] N. Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, “A switching delayed PSO optimized extreme learning machine for short-term load forecasting,” *Neurocomputing*, vol. 240, pp. 175–182, May 2017, doi: 10.1016/j.neucom.2017.01.090.

[18] S. Li, L. Goel, and P. Wang, “An ensemble approach for short-term load forecasting by extreme learning machine,” *Appl Energy*, vol. 170, pp. 22–29, May 2016, doi: 10.1016/j.apenergy.2016.02.114.

[19] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, “Deep learning for estimating building energy consumption,” *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, Jun. 2016, doi: 10.1016/j.segan.2016.02.005.

[20] G. Mohi, U. Din, and A. K. Marnerides, “Short Term Power Load Forecasting Using Deep Neural Networks.” [Online]. Available: http://www.iso-

[21] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, “Deep belief network based electricity load forecasting: An analysis of Macedonian case,” *Energy*, vol. 115, pp. 1688–1700, Nov. 2016, doi: 10.1016/j.energy.2016.07.090.

[22] S. Ryu, J. Noh, and H. Kim, “Deep neural network based demand side short term load forecasting,” *Energies (Basel)*, vol. 10, no. 1, 2017, doi: 10.3390/en10010003.

[23] K. Amarasinghe, D. L. Marino, and M. Manic, “Deep Neural Networks for Energy Load Forecasting.”

[24] L. Saad Saoud, H. Al-Marzouqi, and R. Hussein, “Household Energy Consumption Prediction Using the Stationary Wavelet Transform and Transformers,” *IEEE Access*, vol. 10, pp. 5171–5183, 2022, doi: 10.1109/ACCESS.2022.3140818.

[25] S. Zhang *et al.*, “Architectural Complexity Measures of Recurrent Neural Networks.”