Transfer Learning in Transformer-Based Demand Forecasting For Home Energy Management System

Gargya Gokhale
gargya.gokhale@ugent.be
IDLab, Ghent University — imec
Ghent, Belgium

Jonas Van Gompel
IDLab, Ghent University — imec
Ghent, Belgium

Bert Claessens
IDLab, Ghent University — imec
Beebop.ai
Belgium

Chris Develder
IDLab, Ghent University — imec
Ghent, Belgium

ABSTRACT

Increasingly, homeowners opt for photovoltaic (PV) systems and/or battery storage to minimize their energy bills and maximize renewable energy usage. This has spurred the development of advanced control algorithms that maximally achieve those goals. However, a common challenge faced while developing such controllers is the unavailability of accurate forecasts of household power consumption, especially for shorter time resolutions (15 minutes) and in a data-efficient manner. In this paper, we analyze how transfer learning can help by exploiting data from multiple households to improve a single house's load forecasting. Specifically, we train an advanced forecasting model (a temporal fusion transformer) using data from multiple different households, and then finetune this global model on a new household with limited data (i.e., only a few days). The obtained models are used for forecasting power consumption of the household for the next 24 hours (day-ahead) at a time resolution of 15 minutes, with the intention of using these forecasts in advanced controllers such as Model Predictive Control. We show the benefit of this transfer learning setup versus solely using the individual new household's data, both in terms of (i) forecasting accuracy (~15% MAE reduction) and (ii) control performance (~2% energy cost reduction), using real-world household data.

CCS CONCEPTS

• **Computing methodologies** → **Transfer learning**; • **Applied computing** → **Forecasting**; • **Hardware** → *Smart grid*.

KEYWORDS

Demand Forecasting, Temporal Fusion Transformer, Transfer Learning, Home Energy Management

BuildSys '23, November 15–16, 2023, Istanbul, Turkey

ACM ISBN 979-8-4007-0230-3/23/11…$15.00
https://doi.org/10.1145/3600100.3626635

ACM Reference Format:
Gargya Gokhale, Jonas Van Gompel, Bert Claessens, and Chris Develder. 2023. Transfer Learning in Transformer-Based Demand Forecasting For Home Energy Management System. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), November 15–16, 2023, Istanbul, Turkey.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3600100.3626635

1 INTRODUCTION

The need for a clean and sustainable energy sector has led to significant changes in the modern power grid, including increased integration of renewable energy sources, use of advanced sensors and monitoring devices and growth in electrification [14]. Such changes have been pivotal in the rise of prosumers (i.e., active energy consumers that produce and consume energy) [5]. For our work, we focus on households and residential prosumers operating with a financial objective of reducing energy bills. In most cases, such residential prosumers rely on PV systems for producing electricity and consuming it instantaneously, with the option of injecting the excess to the power grid. Driven by increased volatility in the energy markets [18], we note an increasing shift to storage-based PV systems and more elaborate home energy management systems (HEMS). Supported by increased adoption of sensors, smart meters and other IoT devices, these HEMS make use of advanced control algorithms to identify optimum control strategies for individual prosumers. Model Predictive Control (MPC) has been a dominant control strategy for HEMS, with works such as [7, 9] exploring its applications in diverse settings. In the HEMS context, an MPC entails using a battery model along with forecasts for PV production and household demand to model the household and then using standard optimization algorithms to obtain suitable control policies [4]. However, obtaining accurate forecasts for individual household-level demand and PV production has been a major challenge.

Demand and PV production forecasting is an established research domain with works such as [1, 10] presenting an overview of forecasting techniques specific to household-level demand and PV. Prior works have focused on a range of techniques such as ARIMA-based models [2, 19], hybrid models [3] as well as advanced deep learning methods [13, 20], including recently introduced transformer architectures [8]. While these recent deep learning-based methods show significant improvements in forecasting accuracy, a common problem associated with these methods is their susceptibility to overfitting and the need for large amounts of data to avoid it [17].

Works such as [15, 23] have discussed the use of transfer learning to address this problem. However, these are limited to either hourly time resolutions or work with aggregated forecasts instead of modeling individual households. Only a few previous works such as [13, 21] have utilized a time resolution of less than an hour for their forecasting problems. Forecasting electricity demand of an individual household on a quarter hour basis is an extremely challenging problem, primarily due to the significant influence of user-behavior, which can vary wildly and is difficult to model. Since future power consumption is an important input to optimize HEMS decisions, improving the quality of such forecasts, especially for a quarter hour frequency, can boost the performance of the HEMS, allowing these prosumers to participate more effectively in future energy and consumer-centric markets [6].

The main goal of our work is to investigate how transfer learning methods can help utilize data from multiple different households to improve load forecasts of individual households and develop better control strategies for them. The transfer learning method presented in this work uses data from multiple households to train a global model. This global model can then be finetuned on a new household using only a few days worth of data to obtain a high performing forecasting model for that household. We validate this transfer learning methodology on real-world data obtained from 30 different households and by using the state-of-the-art Temporal Fusion Transformer (TFT) [16] as the forecasting model. Through our simulations, we analyze the day-ahead, quarter hour resolution forecasting performance of our finetuned TFT models as well as the control strategies obtained using these forecasts. Our main contributions can be summarized as:

(1) We propose a transfer learning-based forecasting method using Temporal Fusion Transformers for day-ahead forecasting of individual household-level demand with a quarter hour frequency.
(2) We show that the fine-tuned models require less training data, can generalize to unseen households, and can outperform locally trained TFT models.
(3) Using a simple MPC, we show that such fine-tuned models are effective in obtaining good control policies in home energy management systems.

2 PROBLEM FORMULATION

As discussed in §1, we focus on households and residential prosumers. We consider a single household with a PV system, a small residential battery (5kW, 10kWh), and a dynamic energy tariff. The objective is to develop an energy management system that can minimize the electricity cost for this household by effectively utilizing the battery based on expected demand and PV generation of the household. To develop this energy management system, we formulate a simple MPC-based controller that works with a linear model of the battery and forecasted demand and PV generation profiles. This MPC is designed for a quarter hour control frequency and requires day-ahead forecasts of demand and PV generation profiles.

2.1 MPC Formulation

The objective of the MPC-based energy management system is to minimize the cost of energy consumed by the household. We model the battery using a simple linear model with constant round trip efficiency ($\eta = 90\%$). We assume a dynamic energy tariff with time-varying prices for consuming energy (λ_t^{con}) and injecting energy (λ_t^{inj}) to the grid.[1] Relying on forecasted values of PV generation (P_t^{pv}) and demand (P_t^{con}), the MPC must choose battery actions (u_t) at each time step (t) to minimize the cost of energy consumed over a horizon (T). For our problem, Δt = 15mins. This optimization problem is presented in Eq. (1). Here, E_t refers to the energy state of battery at step t, while $E^{\max}$, $u^{\min}$ and $u^{\max}$ are the energy and power constraints of the battery. P_t^{G} is the power at the meter and we denote power consumed with positive values.

$$
\begin{aligned}
&\min_{u_1,\dots u_T} \sum_{t=1}^{T} c_t \\
&\text{s.t.: } c_t = \begin{cases} \lambda_t^{\text{con}} P_t^{\text{G}} \Delta t & : P_t^{\text{G}} \geq 0 \\ \lambda_t^{\text{inj}} P_t^{\text{G}} \Delta t & : P_t^{\text{G}} < 0 \end{cases} \quad \forall t \\
&P_t^{\text{G}} = P_t^{\text{con}} + P_t^{\text{pv}} + u_t \quad \forall t \\
&E_{t+1} = \begin{cases} E_t + \eta u_t \Delta t & : u_t \geq 0 \\ E_t + \frac{1}{\eta} u_t \Delta t & : u_t < 0 \end{cases} \quad \forall t \\
&0 \leq E_t \leq E^{\max};\ u^{\min} \leq u_t \leq u^{\max} \quad \forall t
\end{aligned}
\tag{1}
$$

Note that, this MPC is formulated as a simple, linear MPC, utilizing a linear model of the battery along with forecasted values of energy consumption and PV generation (i.e., the forecastors are not used in the optimization).

2.2 Demand Forecasting

The MPC problem formulated in Eq. (1), requires forecasted values for PV generation and electricity demand of the household. For this work, we focus on forecasting only the demand (P_t^{con}) and assume exact values for PV production. However, our methods can be extended to PV production forecasting as well. With the intention of using these forecasts for control applications, we model this electrical demand forecasting problem as a univariate stochastic forecasting problem.

3 METHODOLOGY

This section describes our transfer learning-based forecasting approach and provides details related to the forecasting model (Temporal Fusion Transformer) and the experimental setup used for our simulations.

3.1 Temporal Fusion Transformer (TFT)

Building upon the success of the attention mechanism and the transformer architecture in natural language processing domains, the TFT architecture was proposed for multi-horizon time series forecasting [16]. In addition to the self-attention and cross-attention layers used in transformers, TFT uses specialized components, such

[1] We assume that price obtained for injecting energy into the grid is 40% of the price paid for energy consumption. This number can vary depending on the energy contract.

as variable selection networks and gated connections, to effectively encode temporal relationships and obtain an interpretable and accurate forecasting model. These specialized components and the transformer architecture enable TFT to operate on long input series and efficiently capture long-term trends and dependencies in such series [16].

For our work, we use the TFT implementation from Darts [11]. Each input consists of a series of past 672 quarter hours (7 days) of electrical demand for the house, along with temporal features such as hour of day, day of week, etc.[2] The model output is the predictions for the next day's demand (96 steps). We train our TFT models as stochastic predictors using a quantile regression loss function [24]. This leads to trained models that are able to predict when peaks in demand are likely, whereas deterministic forecasters trained with, e.g., mean squared error loss are not able to capture this uncertainty. Further, our quantile forecasts can also be adapted to work with stochastic MPCs in the future.

3.2 Transfer Learning

Transfer learning is used to improve a model from one domain by transferring information from a related domain [22]. In deep learning, transfer learning has been applied to leverage large amounts of data from different sources to pre-train a large model followed by finetuning this model on limited data from the target domain. We follow a similar approach to first train a global TFT model, followed by finetuning this global TFT on individual household's data. We now describe the training and fine-tuning steps for our work.

3.2.1 Global Model. The main idea behind a global model is to use a large set of data to learn good representations related to the commonly occurring patterns in the data. For our case, a global TFT model was trained using 15 months of data from 25 different buildings, amounting to approximately 1M data points. The global TFT model was trained using a quantile regression loss function and had a forecast horizon of 96 steps (24 hours). About 15% of the data was used as validation set for linearly decaying the learning rate and early stopping to avoid overfitting.

3.2.2 Finetuned Models. The global model obtained from the previous step forms the base forecasting model that is to be finetuned for individual households. The finetuning step used a similar training loop as the global model. However, the learning rates and number of epochs in the training loop were significantly reduced. This ensures that during the finetuning phase, the changes to the weights of the TFT are limited and the model does not overfit on the small dataset corresponding to a single household.

3.3 Experiment Setup

The transfer learning approach presented above was implemented on the data obtained as a part of the anonymous BD4NRG research project. This dataset corresponds to 30 real-world households for a period of close to 18 months and contains quarterly measurements of power consumption. The data was preprocessed using standard methods that included filtering null values, aligning time series, and removing outliers. Following this, data from 5 households was held out as test set. The remaining 25 households were used to pretrain the global model. The data from these 25 households was first scaled using a min-max scaler and then split into training and validation sets (85% - 15%). For finetuning, we used data from households in the held-out test set. Finetuning was performed using training data of different sizes (from 14 days to 42 days worth of data) and 1 week worth of validation data. This was followed by testing on 6 weeks of unseen test data. To ensure a common test score across all training sizes, the test set was fixed, and training days corresponded to the n days preceding this test set. The results presented in §4 correspond to this test data.

[2]While works such as [8] recommend using other covariates such as outside air temperature, unavailability of such data led to this design choice.

4 RESULTS AND DISCUSSION

The main idea behind the proposed transfer learning-based forecasting method is to obtain data-efficient forecasts on an individual household-level that can then be used for developing home energy management controllers. To investigate this idea, we first studied the forecasting performance of the finetuned TFT models followed by an evaluation of the control policy obtained when using forecasts from these models. In both cases, the finetuned TFT models were compared with "local" TFT models, i.e., TFT models initialized with random weights and trained solely on data obtained from a single household. This comparison allows us to examine the added value of transfer learning for forecasting performance as well as control performance.

4.1 Forecasting Performance

We tested the forecasting performance of our finetuned TFT models by providing the model with appropriate inputs at the beginning of each test day and obtaining their forecasts for the next 24 hours (to follow the MPC formulated in §2). Each input contained past observations only (i.e., forecasts of the model were never used as inputs) and the process was repeated over the entire test set (6 weeks). The forecasts were compared using Mean Absolute error (MAE). Figure 1 presents the comparison of forecasting performance between our finetuned TFT models and TFT models trained only on local data. The markers depict the mean MAE over the 5-test households and the error bars indicate the standard deviation. It is evident from the figure that the finetuned models are performing significantly better ($\sim$ 15%) than the local TFT models.

4.2 Control Performance

Following the forecasting performance, we now investigate the impact of the forecasts obtained from our finetuned TFT models on the quality of control policies for a simple HEM system. Based on the MPC presented in §2 and using anonymous day-ahead electricity prices, we evaluate the performance of a control strategy that uses a small residential battery (5kW, 10kWh) to reduce the household's energy cost over a 7-day period. Figure 2 shows the mean performance of such an MPC while using forecasts from either our finetuned or local model over the 5-test houses. It can be observed that the MPC with the finetuned model performs slightly better than the local trained models, with an overall cost reduction of $\sim$ 2%. However, this difference in performance between the two controllers is lower as compared to the improvement observed for the forecasting performance, with the performance of finetuned

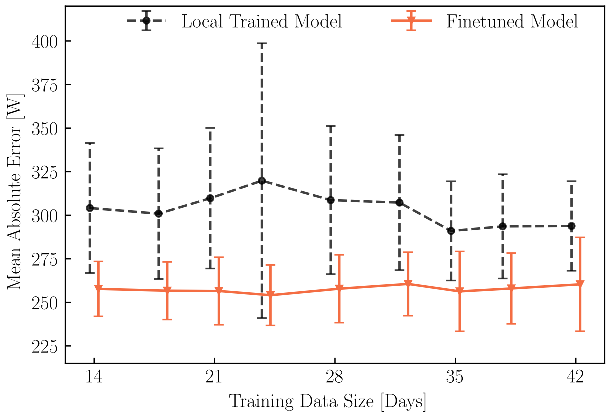

Figure 1: Forecasting Performance of Finetuned TFT model compared with TFT models trained using only local data. The points represent average MAE values over the 5-test households and the error bars represent the standard deviation.

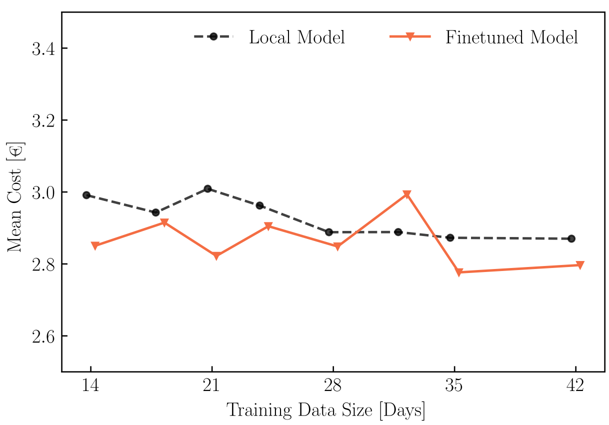

Figure 2: Comparing mean performance of MPC using forecasts from fine-tuned model and a locally trained model over a 1-week period for all 5-test houses.

models even dropping for a few training sizes. This performance gap can be due to simplistic nature of the MPC, the dimensioning of the battery or the small test size (1 week) used for these experiments and will be investigated further in future work.

5 CONCLUSION

Based on the results presented in §4, it is evident that for low training sizes, the finetuned models perform better than the models trained only on individual household data. This supports our hypothesis that using data from different households in a transfer learning-based approach can lead to data-efficient forecasting models that can produce good quality forecasts and can be used in advanced controllers such as MPCs to develop home energy management systems. Following up on these results, we plan to expand the scope of our study in two main research areas. The first one focuses on improving the fine-tuning methodology. This involves creating domain-specific finetuning strategies that can leverage the temporal representations learnt by the global model and combine prior knowledge about the household to fine-tune the forecaster more efficiently. The other research direction will focus on open-source contributions. This involves pretraining global models on a larger set of household data and integrating these global models with platforms such as HuggingFace [12].[3]

ACKNOWLEDGMENTS

This project was performed in collaboration with Centrica Business Solutions (CENTRICA) and we thank the CENTRICA team for their support throughout this study. This work received funding from the European Union's Horizon 2020 research and innovation programme under the projects BD4NRG (grant no. 872613) and BRIGHT (grant no. 957816)

REFERENCES

[1] Kadir Amasyali and Nora M. El-Gohary. 2018. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews* 81 (Jan. 2018), 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095
[2] M. Abdullah Al Amin and Md. Ashraful Hoque. 2019. Comparison of ARIMA and SVM for Short-term Load Forecasting. In *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*. IEEE, "", 1–6. https://doi.org/10.1109/IEMECONX.2019.8877077
[3] Bing Dong, Zhaoxuan Li, SM Mahbobur Rahman, and Rolando Vega. 2016. A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings* 117 (2016), 341–351. https://doi.org/10.1016/j.enbuild.2015.09.033
[4] Ján Drgoňa, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter, Draguna L. Vrabie, and Lieve Helsen. 2020. All you need to know about model predictive control for buildings. *Annual Reviews in Control* 50 (2020), 190–232. https://doi.org/10.1016/j.arcontrol.2020.09.001
[5] EEA 2022. *Institutional members of the TEX Users Group*. EEA. Retrieved Sept 09, 2023 from https://www.eea.europa.eu/publications/the-role-of-prosumers-of
[6] ELIA 2021. *Towards a Consumer Centric and Sustainable Electricity System*. ELIA. Retrieved Sept 09, 2023 from https://www.elia.be/en/news/press-releases/2021/06/20210618-elia-group-publishes-white-paper-on-a-consumer-centric-and-sustainable-electricity-system
[7] Kaitlyn Garifi, Kyri Baker, Behrouz Touri, and Dane Christensen. 2018. Stochastic model predictive control for demand response in a home energy management system. In *2018 IEEE power & energy society general meeting (PESGM)*. IEEE, "", 1–5. https://doi.org/10.1109/PESGM.2018.8586485
[8] Elena Giacomazzi, Felix Haag, and Konstantin Hopf. 2023. Short-Term Electricity Load Forecasting Using the Temporal Fusion Transformer: Effect of Grid Hierarchies and Data Sources. In *Proceedings of the 14th ACM International Conference on Future Energy Systems (e-Energy '23)*. Association for Computing Machinery, New York, NY, USA, 353–360. https://doi.org/10.1145/3575813.3597345
[9] ILR Gomes, MG Ruano, and AE Ruano. 2023. MILP-based model predictive control for home energy management systems: A real case study in Algarve, Portugal. *Energy and Buildings* 281 (2023), 112774. https://doi.org/10.1016/j.enbuild.2023.112774
[10] Stephen Haben, Siddharth Arora, Georgios Giasemidis, Marcus Voss, and Danica Vukadinović Greetham. 2021. Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy* 304 (Dec. 2021), 117798. https://doi.org/10.1016/j.apenergy.2021.117798
[11] Julien Herzen, Francesco LÃssig, Samuele Giuliano Piazzetta, Thomas Neuer, LÃco Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, Maxime Dumonal, Jan KoÅcisz, Dennis Bader, FrÃdÃrick Gusset, Mounir Benheddi, Camila Williamson, Michal Kosinski, Matej Petrik, and GaÃ«l Grosch. 2022. Darts: User-Friendly Modern Machine Learning for Time Series. *Journal of Machine Learning Research* 23, 124 (2022), 1–6.
[12] "HF" 2016. *Hugging Face – The AI Community Building the Future*. Retrieved Sept 09, 2023 from https://huggingface.co/
[13] Haowen Hu, Xin Xia, Yuanlin Luo, Chu Zhang, Muhammad Shahzad Nazir, and Tian Peng. 2022. Development and application of an evolutionary deep learning framework of LSTM based on improved grasshopper optimization algorithm

[3] This is out-of-scope for the work/models presented in this work due to contractual limitations.

for short-term load forecasting. *Journal of Building Engineering* 57 (Oct. 2022), 104975. https://doi.org/10.1016/j.jobe.2022.104975

[14] IEA 2023. *Tracking Clean Energy Progress 2023*. IEA. Retrieved Sept 09, 2023 from https://www.iea.org/reports/tracking-clean-energy-progress-2023

[15] Yuwei Jin, Moses Amoasi Acquah, Mingyu Seo, and Sekyung Han. 2022. Short-term electric load prediction using transfer learning with interval estimate adjustment. *Energy and Buildings* 258 (March 2022), 111846. https://doi.org/10.1016/j.enbuild.2022.111846

[16] Bryan Lim, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. 2020. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. https://doi.org/10.48550/arXiv.1912.09363 arXiv:1912.09363 [cs, stat].

[17] Prajowal Manandhar, Hasan Rafiq, and Edwin Rodriguez-Ubinas. 2023. Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods. *Energy Reports* 9 (Dec. 2023), 2757–2776. https://doi.org/10.1016/j.egyr.2023.01.094

[18] Market Observatory for Energy of the European Commission 2022. *Quarterly Report on European Electricity markets*. Retrieved Sept 09, 2023 from https://energy.ec.europa.eu/system/files/2022-10/Quarterly%20Report%20on%20European%20Electricity%20markets%20Q2%202022_FINAL.pdf

[19] Hongzhan Nie, Guohui Liu, Xiaoman Liu, and Yong Wang. 2012. Hybrid of ARIMA and SVMs for short-term load forecasting. *Energy Procedia* 16 (2012), 1455–1460. https://doi.org/10.1016/j.egypro.2012.01.229

[20] Heng Shi, Minghao Xu, and Ran Li. 2017. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid* 9, 5 (2017), 5271–5280. https://doi.org/10.1109/TSG.2017.2686012

[21] Frederik vom Scheidt, Xinyuan Dong, Andrea Bartos, Philipp Staudt, and Christof Weinhardt. 2021. Probabilistic forecasting of household loads: Effects of distributed energy technologies on forecast quality. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*. ACM, "", 231–238. https://doi.org/10.1145/3447555.3464861

[22] Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. 2016. A survey of transfer learning. *Journal of Big Data* 3, 1 (May 2016), 9. https://doi.org/10.1186/s40537-016-0043-6

[23] Di Wu and Weixuan Lin. 2023. Efficient Residential Electric Load Forecasting via Transfer Learning and Graph Neural Networks. *IEEE Transactions on Smart Grid* 14, 3 (May 2023), 2423–2431. https://doi.org/10.1109/TSG.2022.3208211 Conference Name: IEEE Transactions on Smart Grid.

[24] Keming Yu, Zudi Lu, and Julian Stander. 2003. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society Series D: The Statistician* 52, 3 (2003), 331–350.