Multi-source Transfer Learning in Reinforcement Learning-based Home Battery Controller

Seyed Soroush Karimi Madahi
seyedsoroush.karimimadahi@ugent.be
IDLab, Ghent University – imec
Ghent, Belgium

Toon Van Puyvelde
IDLab, Ghent University – imec
Ghent, Belgium

Gargya Gokhale
IDLab, Ghent University – imec
Ghent, Belgium

Bert Claessens
Beebop
Belgium

Chris Develder
IDLab, Ghent University – imec
Ghent, Belgium

Abstract

On the path to decarbonization, batteries play an important role, as they can help manage uncertainties associated with renewable energy sources such as PV. One of the main challenges of developing a controller for home batteries is its scalability and transferability to other households. In this paper, we propose a multi-source transfer learning framework to scale a pretrained reinforcement learning (RL)-based policy for controlling different unseen households. Our proposed data-driven framework tackles two major disadvantages of vanilla RL methods, i.e., data-intensive training and scalability. In our proposed framework, initially, a global RL agent is trained on multiple source households. Thereafter, the pretrained global agent is finetuned on data from each target household to obtain an individual RL controller for each. We assess the performance of our proposed framework using real data from 100 Belgian households with different load patterns and weather conditions. We benchmark our proposed framework against a rule-based baseline. The results show that our finetuned controllers outperform the baseline rule-based controller by ~8%. Furthermore, compared to agents trained locally from scratch, our finetuned agents require significantly fewer training episodes to learn a good control policy on unseen, target households, validating the scalability of our proposed framework.

CCS Concepts

• **Computing methodologies** → **Transfer learning**; **Reinforcement learning**; • **Hardware** → **Energy generation and storage**.

Keywords

Battery, home energy management system, reinforcement learning, transfer learning

ACM Reference Format:
Seyed Soroush Karimi Madahi, Toon Van Puyvelde, Gargya Gokhale, Bert Claessens, and Chris Develder. 2024. Multi-source Transfer Learning in Reinforcement Learning-based Home Battery Controller. In *The 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '24), November 7–8, 2024, Hangzhou, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3671127.3699534

BuildSys '24, November 7–8, 2024, Hangzhou, China
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0706-3/24/11
https://doi.org/10.1145/3671127.3699534

1 Introduction

Adopting solar PV systems and electrical heating, e.g., heat pumps, can improve environmental sustainability and pave the way for achieving net-zero carbon emissions. Yet, the intermittent nature of PV production implies the need to integrate energy storage, such as batteries. Controlling these batteries furthermore provides flexibility to increase the self-consumption of households or reduce their energy bill [3]. However, developing an optimal controller for batteries in home energy management systems (HEMS) is a challenging task due to the involvement of sequential decision-making under uncertainties caused by PV generation and household consumption.

In literature, different approaches have been used to design controllers for a battery in HEMS, including rule-based control [2], robust optimization [1], model predictive control [4], and reinforcement learning (RL) [14]. Model-based optimization methods have two major drawbacks: first, since these methods require system models, the developed controllers are limited to that specific model and are not easily transferable to other households. Second, because of solving an optimization problem on-the-fly, they might suffer from high computational time, which can make them inefficient for real-time applications. Model-free RL methods overcome these drawbacks, as they do not require system models. They learn a (near-)optimal policy for an environment through a direct interaction with the environment. Nevertheless, data-intensive training and a lack of generalization are two major challenges when using RL in practice [10]. RL methods face the sample efficiency issue, requiring a large amount of observation and exploration for effective training. Moreover, RL agents are capable of finding the optimal policy but specifically for the environment they are trained on. This means that for a new household, a new RL agent must be trained from scratch.

Transfer learning can address the aforementioned challenges of RL by efficiently reusing pretrained RL agents. The main idea of transfer learning is to leverage knowledge from a source domain to improve learning in a target domain that is different but related to the source domain [11]. Using transfer learning for building control and HEMS is an upcoming research area and only a few studies have been conducted on it [5, 8, 10, 13, 15]. However, most previous works focused on either a single source household and a single target household, or on the same source and target domains (for instance, choosing a household as a target that has a similar load profile to the source household).

In this paper, we aim to address this gap in literature and propose a multi-source transfer learning framework that efficiently leverages data from a diverse set of source households to develop RL controllers for new households. First, we train a global RL model on multiple source households with the objective of reducing daily energy cost. Next, we develop a specific agent for each target household by finetuning this pretrained global model on limited data from that target household. We evaluate our proposed framework on 100 diverse Belgian households' data from 5 different locations in Belgium, where 90 households are considered as source households and the rest as target households. Our main contributions are:

1. Propose a multi-source transfer learning framework for developing RL-based controllers for home batteries to minimize daily energy cost;
2. Investigate the effect of using the pretrained global model on training RL agents for unseen target households;
3. Opensource our pretrained global model on HuggingFace[1] to allow other researchers to finetune their specific controllers.

2 Problem Formulation

2.1 MDP Formulation

The home energy management problem we consider is to minimize the daily energy cost of households by controlling the (dis)charging of their home battery. We formulate the problem as a Markov decision process (MDP), which is a mathematical framework for stochastic sequential decision-making problems. The problem is modeled by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma)$, where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ represents the (discrete) action space, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ indicates the reward function, $\mathcal{P} : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$ is the unknown state transition probability distribution, and $\gamma \in (0, 1]$ represents the discount factor [12].

The state at each time step for each household is defined as

$$s_t = (t, \mathrm{SOC}_t, \pi_t, P_t^{\mathrm{PV}}, P_t^{\mathrm{load}}, P_t^{\mathrm{HP}}) \tag{1}$$

where t is the hour of day, SOC_t is the state of charge (SoC) of battery at time t, π_t denotes the electricity price at time t, P_t^{PV} is the PV generation of the household, P_t^{load} represents the non-flexible load consumption, and P_t^{HP} is the heat pump consumption[2].

In this paper, we consider a discrete action space with 5 possible actions, represented as

$$a_t \in \mathcal{A}, \quad \mathcal{A} = \{-P_{\max}, -\frac{P_{\max}}{2}, 0, \frac{P_{\max}}{2}, P_{\max}\} \tag{2}$$

where $P_{\max}$ is the maximum (dis-)charging power of the battery and negative a_t means discharging the battery. In this paper, the decision-making time resolution is 1 hour.

Since the RL agent tries to maximize the cumulative reward, the reward function is formulated as the negative of the energy cost, as shown below.

$$r_t = \begin{cases} -P_t^{\mathrm{agg}} \pi_t^{\mathrm{buy}} & : P_t^{\mathrm{agg}} > 0 \\ -P_t^{\mathrm{agg}} \pi_t^{\mathrm{inj}} & : P_t^{\mathrm{agg}} \leq 0 \end{cases} \tag{3}$$

$$P_t^{\mathrm{agg}} = P_t^{\mathrm{load}} + P_t^{\mathrm{HP}} + a_t - P_t^{\mathrm{PV}} \tag{4}$$

We assume that households are exposed to Belgian dynamic day-ahead prices for their consumption. Also, we consider the injection price to be one-fourth of the day-ahead prices, i.e., $\pi_t^{\mathrm{inj}} = 0.25\pi_t^{\mathrm{buy}}$, which is roughly representative of the injection price in Belgium.

In the MDP framework, a state transition probability function $\mathcal{P}$ models system dynamics. Part of the household dynamics, related to the battery dynamics, can be explicitly formulated. We use a linear model of a 2kW/ 10kWh battery with 90% round-trip efficiency, similar to [6]. However, besides the battery's SoC, the transition function is generally unknown because it depends on stochasticities such as weather, PV generation, and non-flexible household demand. The agent implicitly learns this transition function through interaction with the environment.

2.2 Deep Q Learning

An RL method is used to solve the formulated MDP problem. In this paper, we focus on deep Q learning (DQN) [9], which estimates a Q-function using a deep neural network. The following loss function is minimized to learn the Q-function $Q_\theta(s_t, a_t)$:

$$L = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1}) \sim \mathcal{D}} \left[(r_t + \gamma \max_a Q_{\theta'}(s_{t+1}, a) - Q_\theta(s_t, a_t))^2 \right] \tag{5}$$

For stability in learning, next state-action values in Eq. (5) are calculated based on the target Q-function $Q_{\theta'}(s_t, a_t)$, where $\theta' = \tau\theta + (1 - \tau)\theta'$. Moreover, to avoid overfitting the learned Q function, the neural network is trained on a a mini-batch sampled from an experience replay buffer $\mathcal{D}$. Note that our proposed transfer learning framework is not restricted to DQN and can be easily extended to other RL methods.

3 Proposed Transfer Learning Framework

Transfer learning improves the learning process of a model on a target domain/task by leveraging information from a trained model on a source domain/task. In the context of RL, the task is determined by the reward function (r_t), while the domain consists of the state space ($\mathcal{S}$) and action space ($\mathcal{A}$). We will show one of the main benefits of applying transfer learning to HEMS and building control– For developing a controller using transfer learning for a new household with no historical data or very little data, collecting only a few days' worth of data to finetune the pretrained model could be sufficient. This starkly contrasts with existing RL approaches, where an RL agent, trained from scratch, requires a large amount of training data or environment interactions to learn a good control policy.

3.1 Global Model

Figure 1 illustrates an overview of our proposed framework. We first train a global model on a large amount of data collected from multiple source households (in this study 90 households) to learn a generic control policy. The global model identifies common patterns among households. For example, it learns that batteries need to

[1] https://huggingface.co/soki95/global-model-for-home-battery-controller

[2] We separate the heat pump consumption from load consumption, as heat pumps are controllable assets. We assume that these heat pumps are controlled by their internal logic, reserving RL-based control for heat pumps as future work.

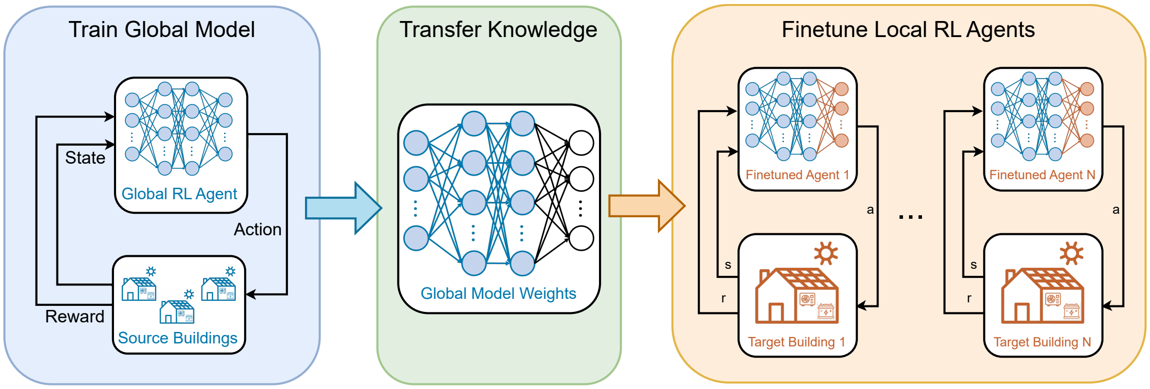

Figure 1: The overview of the proposed transfer learning framework

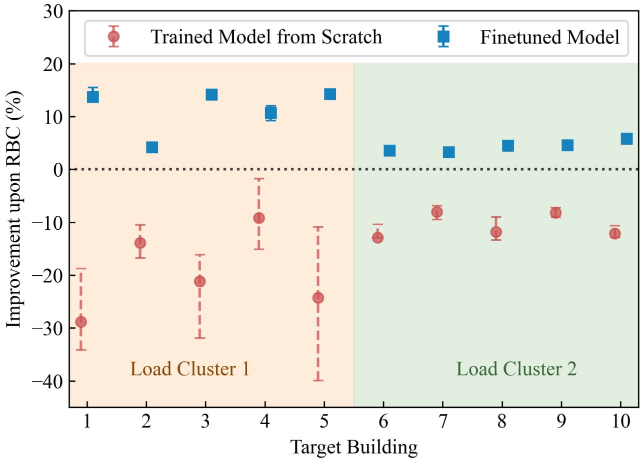

Figure 2: The performance of the finetuned models for the target households on the test days.

be discharged in the evening because there is typically an evening peak in load and price. For this reason, training the global model on a large amount of data can help the agent to better extract such common behaviours.

3.2 Finetuned Models

In this phase, we finetune the global model to obtain a local controller for each target household (in this study 10 target households). These house-specific agents focus on local patterns, such as the typical time when the evening peak in load occurs for each household. The RL training loop in the finetuning phase is similar to that of the global model, with three main changes: (i) For finetuning, we first initialize the local agents using the global model's weight. Further, during finetuning, we freeze most of this model and only tune the weights of the last layer. This ensures that the finetuned agents retain the generic patterns extracted by the global model; (ii) We set a significantly lower learning rate to avoid overfitting on a small dataset; (iii) We finetune the agents for significantly fewer episodes to mimic a situation where only a limited dataset is available.

3.3 Experiment Setup

We selected data from 100 households in Belgium provided by partners in the FlexMyHeat research project[3]. These households belong to different clusters, with their annual energy consumption ranging from 2200 kWh to 8200 kWh. 35 of these households are residential apartments and the other 65 are houses. We designated 10 households as target households and the remaining 90 households as source households to train the global model. These 100 households are located in five different areas to ensure diversity in weather conditions. We used 10 weekdays as a training set, 5 weekdays as a validation set, and 4 unseen weekdays as a test set, all from April 2023. We focused on one month of data to minimize the effects of seasonality on the results. The Q-function and target Q-function were modeled by a fully connected neural network that has two hidden layers with 256 and 128 neurons, respectively.

We benchmarked the trained models against a typical home batteries rule-based controller (RBC), which aims to maximize self-consumption by charging the battery when there is an excess of PV generation and discharging when there is a shortage of PV generation. We trained all global and local models with 10 different seeds to achieve robust results and avoid biased outcomes.

4 Results

To study the efficacy of our proposed framework, we finetuned agents on 10 different, unseen households. Figure 2 shows the performance of these finetuned agents, benchmarked with the RBC and standard RL agents trained from scratch for each of these houses. The markers indicate the mean improvement over the 10 runs, and error bars represent the 25% and 75% quantile values. All of the finetuned and local-from-scratch models were trained for 100 episodes. The figure shows that the finetuned models significantly outperform the models trained from scratch by 19.15% on average. Also, Fig. 2 shows that our proposed framework can scale across different unseen households under scenarios with limited data, as the finetuned agents converge to a good policy and outperform the RBC after only a few episodes (100 episodes). Conversely, we found that the local-from-scratch models need at least 10 000 episodes to surpass the RBC, showing the data-intensive training they require.

[3] https://flexmyheat.ilabt.imec.be/

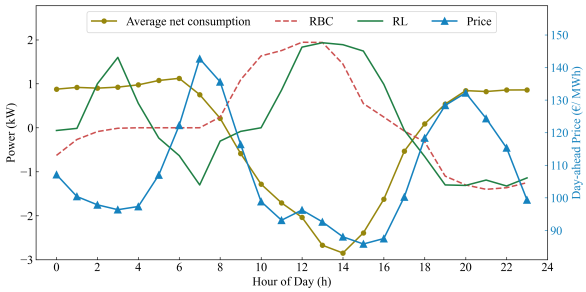

Figure 3: The average learned policy by the global model for the source households on a single test day.

To better understand why the pretrained model jumpstarts the training of the finetuned models, we analyze generic patterns or features learned by the global agent. We trained the global model for 40 000 episodes. The average learned policy of all source households represents the generic behavior learned by the global agent (Fig. 3). The global model charges the batteries from 10:00 to 17:00, when there is high PV generation. In this way, the agent avoids injecting power into the grid at low prices. In the evening, households prevent high-cost electricity consumption from the grid by discharging the batteries for self-consumption. To compensate for the morning demand peak and avoid buying electricity during morning peak hours, the RL agent decides to charge the batteries in the morning when the price is low, which the RBC does not do.

Besides the generic patterns learned by the global model, we also analyzed the feature importance for the learned global model. Figure 4 shows the feature importance obtained using Shapley additive explanations (SHAP) [7]. It reveals that time and PV generation are the most decisive features for the global model. Furthermore, the global model does not pay much attention to the load feature, as it is a house-specific feature and varies among households. It highlights that the global model is able to learn generic, globally common features in the pretraining phase, leaving the local feature learning for the finetuning step.

5 Conclusion

In this paper, we proposed a transfer learning framework to scale an RL-based home battery controller over different unseen target households, by leveraging a pretrained agent on different source households. We validated the performance of our proposed framework using real-world data. Our results show that the finetuned agents learn a high quality control policy for different unseen test buildings. Comparisons with standard RL controllers show that in scenarios with limited data, our finetuned agents can outperform RL controllers trained from scratch, providing improvements of ∼20% over 10 different test buildings. This empirically demonstrates the effectiveness and scalability of our proposed transfer learning framework. Moreover, our results show that the global model jumpstarts the training of the finetuned agents by learning a generic policy, focusing on PV generation and time features. We opensourced this global model, allowing others to use it.

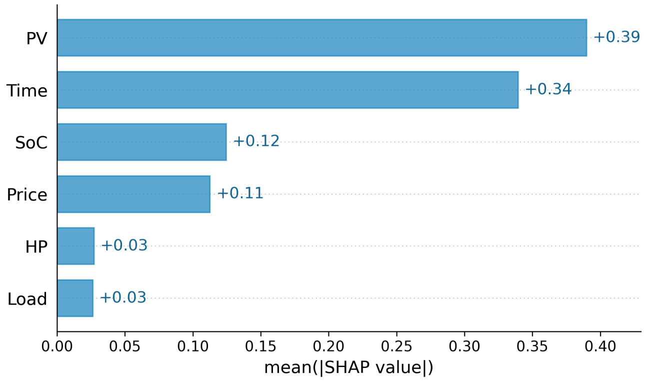

Figure 4: The feature importance for the global model

While the results are promising, we intend to expand this study further, focusing on two main directions. First, we plan to extend our existing transfer learning framework to handle cases where the source and target reward functions (tasks) are different. For instance, how to finetune a global model, trained on source households to minimize the energy cost, to achieve peak shaving in target households. A second direction we intend to focus on is developing an offline framework for transfer learning where all global and local agents will be solely trained on collected historical data, without any interaction with simulation or real-world environments.

Acknowledgments

This research has received funding from the Energy Transition Fund (FOD Economy) via the FlexMyHeat project and by Daikin Europe NV under the PhD Framework Agreement with Ghent University.

References

[1] Jonas Engels, Bert Claessens, and Geert Deconinck. 2017. Combined stochastic optimization of frequency control and self-consumption with a battery. *IEEE Transactions on Smart Grid* 10, 2 (2017), 1971–1981.

[2] Gargya Gokhale, Bert Claessens, and Chris Develder. 2024. Explainable Home Energy Management Systems based on Reinforcement Learning using Differentiable Decision Trees. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 687–691.

[3] Gargya Gokhale, Seyed Soroush Karimi Madahi, Bert Claessens, and Chris Develder. 2024. Distill2Explain: Differentiable decision trees for explainable reinforcement learning in energy application controllers. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 55–64.

[4] Gargya Gokhale, Jonas Van Gompel, Bert Claessens, and Chris Develder. 2023. Transfer Learning in Transformer-Based Demand Forecasting For Home Energy Management System. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 458–462.

[5] Fangli Hou, Jack CP Cheng, Helen HL Kwok, and Jun Ma. 2024. Multi-source transfer learning method for enhancing the deployment of deep reinforcement learning in multi-zone building HVAC control. *Energy and Buildings* (2024), 114696.

[6] Seyed soroush Karimi madahi, Gargya Gokhale, Marie-Sophie Verwee, Bert Claessens, and Chris Develder. 2024. Control Policy Correction Framework for Reinforcement Learning-based Energy Arbitrage Strategies. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 123–133.

[7] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.

[8] Brida V Mbuwir, Kaveh Paridari, Fred Spiessens, Lars Nordström, and Geert Deconinck. 2020. Transfer learning for operational planning of batteries in commercial buildings. In *2020 IEEE International Conference on Communications,*

Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE, 1–6.

[9] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *nature* 518, 7540 (2015), 529–533.

[10] Kingsley Nweye, Siva Sankaranarayanan, and Zoltan Nagy. 2023. MERLIN: Multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities. *Applied Energy* 346 (2023), 121323.

[11] Thijs Peirelinck, Hussain Kazmi, Brida V Mbuwir, Chris Hermans, Fred Spiessens, Johan Suykens, and Geert Deconinck. 2022. Transfer learning in demand response: A review of algorithms for data-efficient modelling and control. *Energy and AI* 7 (2022), 100126.

[12] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT Press.

[13] Shichao Xu, Yixuan Wang, Yanzhi Wang, Zheng O'Neill, and Qi Zhu. 2020. One for many: Transfer learning for building hvac control. In *Proceedings of the 7th ACM international conference on systems for energy-efficient buildings, cities, and transportation*. 230–239.

[14] Yang Xu, Weijun Gao, Yanxue Li, and Fu Xiao. 2023. Operational optimization for the grid-connected residential photovoltaic-battery system using model-based reinforcement learning. *Journal of Building Engineering* 73 (2023), 106774.

[15] Xiangyu Zhang, Xin Jin, Charles Tripp, David J Biagioni, Peter Graf, and Huaiguang Jiang. 2020. Transferable reinforcement learning for smart homes. In *Proceedings of the 1st international workshop on reinforcement learning for energy management in buildings & cities*. 43–47.