RESEARCH ARTICLE

Optimal price-based control of heterogeneous thermostatically controlled loads under uncertainty using LSTM networks and genetic algorithms [version 1; peer review: 2 approved with reservations, 1 not approved]

Taha Abdelhalim Nakabi , Pekka Toivanen

School of Computing, University of Eastern Finland, Kuopio, 70211, Finland

V1 First published: 10 Sep 2019, 8:1619
<https://doi.org/10.12688/f1000research.20421.1>
 Latest published: 10 Sep 2019, 8:1619
<https://doi.org/10.12688/f1000research.20421.1>

Abstract

In this paper, we consider the problem of thermostatically controlled load (TCL) control through dynamic electricity prices, under partial observability of the environment and uncertainty of the control response. The problem is formulated as a Markov decision process where an agent must find a near-optimal pricing scheme using partial observations of the state and action. We propose a long-short-term memory (LSTM) network to learn the individual behaviors of TCL units. We use the aggregated information to predict the response of the TCL cluster to a pricing policy. We use this prediction model in a genetic algorithm to find the best prices in terms of profit maximization in an energy arbitrage operation. The simulation results show that the proposed method offers a profit equal to 96% of the theoretical optimal solution.

Keywords

Artificial intelligence, Artificial neural networks, Customer behavior learning, Demand response programs, Energy arbitrage, LSTM, Partial observability, Price elasticity of demand, Profit maximization, Smart grid, thermostatically controlled loads.

This article is included in the **Artificial Intelligence and Machine Learning** gateway.

Open Peer Review

Approval Status ? ? X

	1	2	3
version 1	?	?	X
10 Sep 2019	view	view	view

1. **Matsukawa Shun**, Gifu University, Gifu, Japan
2. **Chiou-Jye Huang** , Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China
3. **Shanti Swarup** , Indian Institute of Technology Madras, Chennai, India

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Mathematical, Physical, and Computational Sciences** collection.

Corresponding author: Taha Abdelhalim Nakabi (tahanak@uef.fi)

Author roles: **Nakabi TA:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Toivanen P:** Funding Acquisition, Supervision

Competing interests: No competing interests were disclosed.

Grant information: Jenny and Antti Wihuri Foundation.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Nakabi TA and Toivanen P. This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Nakabi TA and Toivanen P. **Optimal price-based control of heterogeneous thermostatically controlled loads under uncertainty using LSTM networks and genetic algorithms [version 1; peer review: 2 approved with reservations, 1 not approved]** F1000Research 2019, **8**:1619 <https://doi.org/10.12688/f1000research.20421.1>

First published: 10 Sep 2019, **8**:1619 <https://doi.org/10.12688/f1000research.20421.1>

Abbreviations

DR	Demand Response
GA	Genetic Algorithm
LSTM	Long-Short Term Memory
MDP	Markov Decision Process
TCL	Thermostatically controlled load

Indices

n	Index for TCL units $n = 1, 2, \dots, 30$
t	Index for time step, $t = 1, 2, \dots, 24$

Parameters

f	Transition function
g	State approximation function
H	Control horizon
L_{max}	Maximum load capacity ¹
L_{thr}	Load threshold
N	Number of TCL units to control
PN	Population size
R_{max}	Revenue cap
U	Action space
W	Random process space
X	State space

Variables

$\underline{C}$	Candidate state vector in LSTM network
C_t	Cost function at time t
ΔT_t	Gap between the outdoor and indoor temperatures [$^{\circ}\text{C}$]
h	Control policy
$\underline{h}_t$	Hidden state vector of LSTM network
$I_{n,t}$	Input matrix of LSTM network
P_t	Selling electricity price at time t [€ cent/kW]
P_t	Wholesale electricity price at time t [€ cent/kW]
$P_{t,max}$	Maximum selling price at time t [€ cent/kW]
$P_{t,min}$	Minimum selling price at time t [€ cent/kW]
P_w	Probability distribution
p	Control action reward
T_t	Temperature at time t [$^{\circ}\text{C}$]
u_t	Control action at time t
x_t	State at time t

¹This work was supported by The Jenny and Antti Wihuri Foundation, FINLAND.

Introduction

In a power network relying on distributed and renewable energy resources, the exploration of new sources of flexibility is a key factor for its stability. Given the intermittent nature of renewable energy resources, it is challenging to maintain the power balance under normal operating conditions in a grid with deep penetration of these resources. Therefore, more integration of renewable resources increases the need for ancillary services such as regulation reserve and load following requirements¹. However, using traditional fossil fuel generators to provide these reserves will decrease the net carbon benefit from renewables, weaken generation efficiency and will be economically untenable. Alternatively, demand-side resources can play a key role in supplying the regulation service needed for deep renewable integration with zero-emission operations. Demand-side resources such as thermostatically controlled loads (TCLs), electric vehicles and strategic storage can contribute to ancillary services by acting as a source of flexibility to the grid. Unlike the traditional demand-side management programs, such as peak load shaving and emergency load management, the exploration of higher flexibility from the above-mentioned loads has a big potential in offering more lucrative and faster ancillary services. The potential of these sources of flexibility is reflected on the energy market. Electricity prices fluctuate according to the availability and demand of energy. This can open considerable opportunities for energy arbitrage².

A significant potential for provision of flexibility resides in TCLs such as air conditioners ACs, heat pumps, water heaters, and refrigerators. TCLs represent a high percentage of the total electricity consumption^{3,4}. The nature of TCLs permits them to act as a thermal storage which makes it possible to adjust their electricity consumption while maintaining the temperature requirements and the comfort level of the end user. The idea of TCL flexibility relies on the principle that the temperature constraints specified by the users, can be fulfilled by different power trajectories. Finding the optimal trajectory that provides the required flexibility and high lucrative ancillary service is the subject of several studies⁵⁻⁷. However, this problem requires real-time information about the state of TCLs, their envelope temperature and their behavior in response to temperature dynamics. In most of the cases, this information is only partially available and requires qualitative or quantitative models to estimate it. It is also possible to use model-free approaches to solve the problem of uncertainty and find near-optimal power trajectories².

The optimal power trajectory for a cluster of TCLs is then translated to individual or aggregated control signals using a variety of control methods. Control methods can be categorized into intrusive forms, including direct and indirect control, and non-intrusive form using price proxies. The direct intrusive form of control consists of directly controlling the on/off states of the TCLs, the indirect intrusive form consists of controlling the parameters of TCLs, such as the temperature set points and the switch cycles and the non-intrusive form of control uses dynamic prices to steer the consumption of TCLs relying on price-based demand response programs. The intrusive

form requires an aggregator contracting with each TCL unit holder for taking control of their TCLs with the condition that their temperature constraints will be respected throughout the control period. The non-intrusive approach relies on the end user's involvement and response to a given control signal in return of a certain incentive or special pricing. The users' response to these signals can also be an automatic response to electricity prices throughout the day using home energy management systems or embedded TCL controllers⁸.

Intrusive control of TCLs has a big potential in offering a wide range of flexibility and market opportunities for the aggregators. It offers a faster response to control signals and permits the design of a more reliable energy arbitrage strategy compared to non-intrusive control through price proxies. However, the implementation of the technological requirements for an intrusive control on a large scale can be challenging due to its high financial requirements. Additionally, the question of whether the consumers are ready to give up the control of their TCLs to an external party can also be a barrier for the implementation of these programs. According to ⁹, the integration of end users in the demand response (DR) programs is a key factor for its success. Several smart grid projects were analyzed from this perspective and the conclusions suggest that more attention should be given to the domestication of these technologies and their adaptation with the users' experience considering their social dimensions such as individual behavior, education, and income level^{9,10,11}. It is therefore necessary to include all these factors in the design of a DR program. Non-intrusive control, on the other hand, has fewer constraints regarding the users' comfort and data privacy. It makes the end user feel included in the decision making of the grid and involved in the energy management. This discussion can serve as a benchmark when making the choice of the control strategy and the implementation of a large-scale DR program.

In our paper, we choose to implement a non-intrusive control using dynamic electricity prices. We first formulate the problem as a Markov decision process (MDP)¹², where the policy consists of a sequence of electricity prices. The agent is assumed to have no prior knowledge or data about the state of TCL units except their real time power consumption. The idea is to use data-driven models that can learn the consumption patterns of each individual TCL unit and their response to temperatures and prices. We use a long-short-term memory (LSTM) neural network architecture to learn individual TCL units' behaviors as in ¹³. This method can overcome the problem of uncertainty and the diversity of power consumption preferences in response to varying prices. The aggregator uses these models to simulate the aggregate response TCLs to different pricing schemes during a certain control horizon. An optimization algorithm is then applied to find the best pricing strategy given an objective function. When controlling a cluster of TCLs, different objective functions are considered in the literature, such as tracking a balancing signal⁷ or energy arbitrage⁵. In this work we adopt an energy arbitrage objective function, where we maximize the profit of an aggregator that buys electricity from the wholesale market and sells it in the retail market to end users with TCL units. A genetic algorithm is implemented to find the best pricing solution of the aggregate TCLs.

Related work and contributions

The literature contains extensive research concerning TCL control and their flexibility potential.

TCL control approaches

Most early studies, as well as current work, focus on direct intrusive control methods and frameworks. Early work that tackled aggregated modeling of TCLs can be found in ¹⁴ and ¹⁵. The solution computation and controller design of these approaches is considerably difficult, which represents a drawback for these approaches. These issues were mitigated in more recent works^{5,7,16} using a different class of linear population-bin transition models based on Markov chains. Other approaches have proposed time-varying battery models with dissipation such as ¹⁷ or without dissipation as in ¹⁸. These approaches were used to compute near-optimal control trajectories with a reduced computational cost. Although optimal pricing for demand side management has been thoroughly studied in the literature^{19–21}, the price-based control of TCLs remains only briefly addressed in the literature. In ²², the operating reserve capacity of aggregated heterogeneous TCLs was evaluated using a TCL model that takes into consideration consumer behavior. The price-based approach was also addressed from the consumer perspective in ²³. The objective of the proposed method was mainly to find the optimal set point change in response to electricity prices in order to minimize the increases in the electricity bill due to dynamic pricing. The power gain from this control scheme was then used for load following supply. Another approach was proposed to find the equilibrium between the electricity prices and the users' comfort. Using a Stackelberg game approach, authors in ²⁴ presented a unique Stackelberg equilibrium that maximizes the utility function and minimizes dissatisfaction cost of TCLs users. A similar approach was proposed in ²⁵ and ²⁶ using a mean-field game approach to find the best pricing scheme considering TCLs as price-responsive rational agents.

Deep learning-based models for TCL control with partial observability

Deep learning and other machine learning methods are largely applied in DR programs²⁷. The implementation of a TCL cluster control program faces the problem of uncertainty and heterogeneity of the TCL units' behaviors in response to control prices. Consequently, many researchers were interested in using machine learning models that can learn aggregate or individual behavior of TCL units under partial observability. A model-free reinforcement learning was early proposed in ²⁸ for TCL control that gives similar results as model predictive approaches. Reinforcement learning approaches were also used in²⁹ to control domestic water buffers according to a local photovoltaic production for the maximization of self-consumption. More recently, the success of deep reinforcement learning approaches has inspired more researchers to tackle the problem of direct TCL control using deep reinforcement learning. Authors in ^{30–33} have used different deep neural architectures for automatic estimation of the TCLs' state's features in a batch reinforcement learning model. The same authors have later provided a comparison of the different architectures in ^{33,34}. The LSTM architecture has outperformed the other deep neural network architectures. These works focused only on deep Q-learning, which is based on the estimation of a quality

function for every potential action before performing the optimization. In 35 Deep policy gradient method was explored along with deep Q-learning for an on-line energy optimization of the buildings.

Contributions

Following the above-mentioned literature and the success of LSTM networks in mitigating the problem of partial state information and solving long-term dependency problem^{13,33,34}, we propose a two-step pricing optimization method for the exploration of TCL flexibility in energy arbitrage. This paper addresses the need for new non-intrusive TCL control methods via electricity prices proxies, so far lacking in the scientific literature. The proposed method relies on LSTM networks learning individual TCL unit behavior and the prediction of individual responses to electricity prices. The individual predictions are aggregated to form an overall prediction model. This model is used in a genetic algorithm (GA)-based optimization algorithm to maximize a retailer's profit considering grid and energy cost constraints. To the best of the authors' knowledge, this is the first work that uses LSTM networks in a non-intrusive TCL control problem based on electricity prices within a DR program. The main contributions of this paper are the following:

- An MDP formulation of the price control problem where the policy is the set of electricity prices during a control horizon.
- An LSTM network for learning the individual behavior of TCL units in response of electricity prices and temperatures.
- An aggregation of individual TCL units' behaviors, in response to prices, to derive a global estimation of the potential response of the TCL units cluster.
- A genetic algorithm that uses the aggregated information from the LSTM networks to optimize the lucrative benefits from an energy arbitrage operation.

Problem formulation

We consider a cluster of residential households powered by electricity from the same retailer or utility company. The households are equipped with smart meters and TCLs that can react to electricity prices and indoor temperatures. The retailer implements a price-based DR program that announces electricity prices for a certain time horizon in such a way that maximizes an objective function. The optimization is based on an estimated information about the responsiveness to electricity prices and temperatures. Before discussing the pricing optimization approach, we formulate the problem as an MDP¹². An MDP is defined by its state space X , its action space U , and its transition function f , which defines the dynamics between the current state $x_t \in X$ and the next step x_{t+1} under a control action $u_t \in U$ and subject to a random process $w \in W$ with a probability distribution $p_w(\cdot, x_t)$. The transition equation is defined as follows:

$$x_{t+1} = f(x_t, u_t, w_t) \quad (1)$$

The objective of this process will be to find a policy $h: X \rightarrow U$ that minimizes or maximizes a cost function or a reward function throughout the control horizon starting from a state x_1 denoted by:

$$R_h(x_1) = E(\sum_t \rho(x_t, h_t, w_t)) \quad (2)$$

where ρ is the reward or the cost of each time step k given an action h_t . Unlike the classic Q-iteration methods, the policy is characterized directly by sum of rewards during a time horizon H . The optimization is performed on the set of actions during the time horizon H and the fitness function is the cost function R_h of the policy h . For each policy h , a corresponding sequence of states is estimated implicitly by the forecasting model.

State and control action description

The agent is only able to measure a partial observation of the true state i.e. no information about the indoor temperatures, resulting in a partially observable Markov decision problem. The observable state space X consists of two variables: the outside temperature, and the electric load:

$$x_t = (L_t, T_t) \quad (3)$$

Since the observable state space only includes part of the true state, it is not possible to directly model future state transitions. Yet this remains convenient when following the results from 13 that we can predict the next step electric load L_{t+1} using the information of outdoor temperature T_t , the electric load L_t and the electricity price P_{t+1} . The state is extended with sequences of past observations of states and actions, which results in a non-Markovian state.

For each TCL, the electric load is approximated by:

$$L_{t+1} \sim g(L_t, T_t, P_{t+1}) \quad (4)$$

We assume that the outside temperatures' forecasts are available for every future timestep in the control horizon.

The control action u_t consists of the electricity price that the retailer announces for each time step of the control horizon. As mentioned earlier, even though the retailer is not controlling the TCLs directly, we assume that the TCLs react directly to electricity prices. Therefore, the electricity price controls the state by influencing the amount of energy consumed during a timestep t . The next state is then defined by:

$$\begin{aligned} x_{t+1} \\ = f(x_t, P_t, w_t) \sim (g(L_t, T_t, P_{t+1}), T_{t+1}). \end{aligned} \quad (5)$$

Objective function

According to the existing literature, the control of TCLs clusters can be performed considering different objective functions. For instance, the objective can be tracking a balancing signal or energy arbitrage. In this work we consider an energy arbitrage problem where a retailer is trying to maximize

their profit. However, the framework and methods presented here might as well be applied to different objective functions. We consider the profit as the difference between the revenue and the cost function. We assume that the cost function $C_i(L_i)$ is convex increasing in L_i for each timestep as formulated in 36.

$$C_i(L_i) = qL_i^2 + p_iL_i + c \quad (6)$$

where, $q > 0$ is a constant, $p_i > 0$ is the electricity price in the wholesale market and $c > 0$ is a fixed cost.

In order to avoid overload during peak times, we introduce a maximum load capacity of the power network, denoted $L_{t,max}$ at each timestep. Therefore, we have the following constraint:

$$L_t = \sum_n L_{n,t} \leq L_{t,max}, \quad \forall t \in H \quad (7)$$

The revenue is the bill that customers would pay for using the energy during the time window H :

$$R = \sum_{t=0}^H (L_t * P_t) \quad (8)$$

Usually, there exists a total revenue cap, denoted as R_{max} , for the retailer. Therefore, we need to add the revenue constraint to improve the acceptability of the retailer's pricing strategies, i.e., without such a constraint, the retail prices will keep going up to a level which is against energy regulations as well as financially unacceptable to the customers. As a result, we have the following constraint:

$$R < R_{max} \quad (9)$$

Moreover, for each timestep $t \in H$, we define the minimum and maximum price that the retailer (utility company) can offer $P_{t,min}$ and $P_{t,max}$, we have:

$$P_{t,min} \leq P_t \leq P_{t,max}, \quad \forall t \in H \quad (10)$$

$P_{t,min}$ and $P_{t,max}$ are usually designed based on historical prices, market competition, customers' acceptability, and the wholesale price. It is reasonable to assume that the price the retailers can offer is greater than the wholesale price for each hour, and there exists a price cap for the retail prices due to retail market competition.

Finally, the control problem defined the optimization of the price vector P , during the time horizon H , can be modeled as follows:

$$\max_p \{R - \sum_{t=0}^H C_i(L_i)\} \quad (11)$$

subject to constraints:

$$R < R_{max} \quad (12)$$

$$L_t \leq L_{max} \quad (13)$$

$$P_{t,max} \leq P_t \leq P_{t,max}, \quad \forall t \in H \quad (14)$$

Methods and implementation

Given the partial observability of this problem, the methods proposed in this paper are nondeterministic. An LSTM network is used to estimate the next states given an initial state and a pricing policy. The method consists of learning the individual behavior of each TCL agent n using an LSTM method as illustrated in 13. The N estimation models will predict the reaction $L_{n,t+1}$ of each TCL to a state x and a pricing action P_t . The overall estimated load L_t is the sum of all the load predictions as in (7). Given this estimation model, we apply a genetic algorithm to find the best pricing policy.

LSTM networks for state estimation

LSTM networks are recurrent neural networks that consist of memory blocks. These memory blocks replace the summation units in the hidden layers in a standard recurrent neural network. The input vector and the hidden state vector are passed through the forget gate to determine the keeping rate of the cell state components. The same vector is passed through the input gate to determine how much of the new state candidate $\underline{C}$ can pass to the new cell state. Finally, the output gate will decide how much of the transformed state cell vector can be passed to the next hidden state vector $\underline{h}_t$. Following 13, the proposed LSTM network consists of multiple layers of LSTM cells followed by a fully connected layer as illustrated in Figure 1. In the case of our model, the input $I_{n,t}$ is a 2×3 matrix that consists of the electric loads, the temperatures and the electricity prices as follows:

$$I_{n,t} = \begin{pmatrix} L_{n,t-1}, T_{t-1}, P_t \\ L_{n,t}, T_t, P_{t+1} \end{pmatrix} \quad (15)$$

The LSTM network recurrently uses the historical information of loads, temperatures and prices to predict electric load for an individual TCL n , in the next timestep. The aggregation of these predictions gives an approximation of g function mentioned in the previous section.

Initially, for each TCL agent $n \in N$ we train an LSTM network based on the historical reactions of these TCLs to prices and temperatures. We assume that a DR program is implemented during a long period, enough to collect a sufficient amount of data related to the reactions of TCL agents to prices and temperatures.

Genetic algorithms for price optimization

Due to the discontinuous nature of the objective function and the complicated dependency between the function electric load L and the electricity prices P , the conventional nonlinear optimization methods are not usable for this problem. Therefore, GA-based optimization algorithms are more suited for this problem³⁷. The proposed GA algorithm uses rank selection and value encoding³⁸. Each chromosome represents a pricing policy

Figure 1. LSTM Network for TCLs load prediction. The model uses the information about temperatures, loads and price in the previous timesteps to predict the load $L(t)$. Since this is a regressions problem, the fully connected layer uses a linear activation function.

P and consists of a vector of size H . We use uniform crossover³⁹ and non-uniform mutation⁴⁰. The constraints are handled by the approach proposed in 41.

The proposed GA-based optimization algorithms for TCL pricing control are given in Algorithm 1 and Algorithm 2.

Algorithm 1. GA-based optimization algorithm for TCL pricing control.

- 1: Population Initialization, i.e., generating a population of PN chromosomes randomly; each chromosome denotes a pricing policy for the next time horizon H .
- 2: **for** $i=1$ to PN **do**
- 3: Concatenate the price vector to the temperature forecasts of the next time horizon.
- 4: **for** each TCL agent n in N **do**:
- 5: Use LSTM network iteratively to predict $(L_{n,t'})_{t' \in H}$ using Algorithm 2.
- 6: **end for**
- 7: Calculate L_p , $C_p(L_p) \forall t \in H$, and R
- 8: Check the feasibility of policy P regarding the constraints. Handle the invalid individuals by the approach proposed in []. Then calculate the fitness value of policy P .
- 9: **end for**
- 10: Create a new generation of chromosomes by using the selection, crossover, and mutation operations of the GA.
- 11: Repeat steps 2–11 until the stopping condition is reached.
- 12: Announce the best price vector via the two-way communication infrastructure at the beginning of the control horizon.

Algorithm 2. Individual TCL load prediction using LSTM network.

- 1: Build the initial input matrix $I_{n,0}$ using the initial values of prices, loads and temperatures.
- 2: **for** $t=0$ to H **do**
- 3: Use the input matrix $I_{n,t}$ to predict $L_{n,t+1}$
- 4: Concatenate L , T and P with the last line of the input matrix $I_{n,t}$ to build the next input matrix:

$$I_{n,t+1} = \begin{pmatrix} L_{n,t}, T_{n,t}, P_{n,t+1} \\ L_{n,t+1}, T_{n,t+1}, P_{n,t+2} \end{pmatrix}$$

- 5: **end for**
- 6: **return** $(L_{n,t'})_{t' \in H}$

In Algorithm 1, we initialize a population of NP pricing policies at step 1. For each policy P we perform steps 2–6 to evaluate the fitness function and the feasibility for each policy. The evaluation of policies is performed using LSTM sequence prediction presented in Algorithm 2. The best policies are selected, and a new generation is created using crossover and mutation operations in step 10. This process is repeated until a stopping condition or maximum number of iterations is reached. At the end of the optimization process, the best pricing policy is selected, and prices are announced to TCL agents via two-way communications technology. After each control episode, the LSTM learning models are updated according to the new data collected from the actual response to the implemented electricity prices.

Results

In this section we evaluate the functionality of the proposed pricing control methods. A set of numerical experiments were

performed on a simulation scenario comprising a population of 30 TCLs exposed to dynamic electricity prices during a period where the outdoor temperatures change significantly. The thermal inertia of each TCL allows the electric demand to be shifted towards lower price moments. The TCL agents determine the amount of electricity to be consumed at each timestep according to the indoor temperature and the electricity prices. The objective of TCL agents is to maintain a reasonable comfort level while minimizing the electricity bill. Therefore, the different TCL agents have different reactions given a set of prices and temperatures depending on individual user's preferences and buildings' characteristics. We define a control timestep of 1 hour and a control horizon of 6 hours. The choice of the control horizon is justified by the limited ability of LSTM to predict large sequences of the future electric loads. The control horizon is chosen in a way that minimizes the number of times the retailer runs the control algorithms and announces the prices, while keeping a good accuracy of the LSTM predictions.

Simulation data

Following 13 the simulation data is generated using two fuzzy logic systems with the following assumptions:

- The TCL agents are reacting to indoor temperatures and electricity prices.
- The difference between the outdoor and indoor temperature ΔT depends on the building characteristics and the amount of energy spent in heating/cooling in previous timesteps.

TCL agents are operating during the day to maintain a comfortable temperature of the space while taking into consideration the electricity price in a given hour. Fuzzy logic is used in this problem because it can model non-qualitative concepts like "hot temperature" or "low price". The combination of the two fuzzy logic systems delivers the load $L_{n,t+1}$ using the outdoor temperature T_t and the electricity price P_{t+1} . The simulation is performed with different parameters to generate diverse data for 30 TCL agents. The temperature and price data used for the simulation are taken respectively from the *Kaisaniemi* observation station in Helsinki, available online in 42, and *Elspot DA* electricity prices in Finland⁴³ for the period between 1st January 2017 and the 7th September 2018. The generated dataset consists of 14,734 data points for each TCL agent.

LSTM networks results

The data generated from the above-mentioned simulations is used to train the LSTM networks to learn the behavior of each individual TCL agent. The hyperparameters and structure of the LSTM networks are chosen according to the results of 13 and summarized in Table 1.

The results are evaluated using validation data generated from the same simulations. Figure 2a illustrates the learning results for three TCL agents during different time periods with different temperatures and prices. Figure 2b illustrates the comparison between the real and predicted average power consumption of the 30 TCL agents cluster. The power curves

Table 1. Results of LSTM model hyperparameters optimization.

Sequence length	2
LSTM cell size	30
LSTM cells	2
Dropout	0.2
Activation	'tanh'
Recurrent activation	'selu'
Optimizer	'rmsprop'

show that the TCL agents' responses to prices and temperatures are slightly different. In general, the power consumption is high when the temperatures and electricity prices are low and vice-versa. The comparison between the true load curves and the predicted load curves show a very small prediction error per hour in most cases. The true and predicted load curves have similar shapes and significant resemblances. The peaks and valleys are also predicted accurately in most of the cases, which gives a valuable insight for demand side management.

GA Optimization results

We run the GA optimization algorithm on a population of size 100 for 100 iterations. The parameters used for the optimization are summarized in Table 2. The optimization process is graphically presented in Figure 3. The learning process is measured by the fitness of the best individual in the population at each iteration. Figure 4 illustrates the results of the best pricing solutions for one day. Figure 4a is an illustration of the electricity prices fluctuations during the 24 hours. Figure 4b shows a comparison between the power consumption of the whole cluster under original prices and the power consumption under optimized prices. Figure 4c presents the revenue and profit that the retailer would make under original and optimized prices. Figure 4d presents daily bill of each user of the cluster under original and optimized prices.

The results show a general increase in prices throughout the day. However, this increase didn't result in an increase in the daily electricity bills. Most of customers will be paying a slightly lower amount per day. This is a consequence of upper limit constraint on the revenue described in (12). The overall consumption of electricity was decreased comparing to the original pricing scheme which gives a good idea about the potential energy saving that an optimal pricing strategy can offer.

Comparison with a theoretical benchmark

In order to validate the performance of the proposed algorithm, we consider a case where we have a full access to TCL units' behavior, i.e. the exact electricity consumption of each TCL unit given temperatures and prices at each timestep. The optimization is performed with direct access to the simulation model described above, which provides full observability and perfect information about the TCLs. This theoretical setup can

Figure 2. LSTM Learning results. (a) Power consumption of different TCL agents in response to electricity prices and outdoor temperatures. (b) Average real and predicted power consumption of the cluster surrounded by an envelope containing 9% of the power consumption profiles for different days.

Table 2. Optimization parameters.

PN	100
L_{max}	75.0 kWh
q	0.01 €cents/[kWh] ²
c	1.0 €cents
$P_{t,min}$	p_t
$P_{t,max}$	$2 * p_t$
R_{max}	$N * H * 5.5$ €cents

serve as a benchmark of our method. It can be seen as an upper limit on the profit possibly made by the aggregator without violating the constraints.

The results illustrated in Figure 5a–d, show that the proposed methods have performed very similarly to the benchmark. The hourly prices in Figure 5a, are only slightly shifted from the benchmark prices during most of the day. The difference is only significant in 2 to 3 points. The same observation can be

made for the revenues and profits in Figure 5b and electricity consumption in Figure 5c. The comparison of daily bills under optimized prices and benchmark prices in Figure 5d shows a slight rise in the electricity bill in the benchmark model for most customers. This can be explained by the slight increase in prices illustrated in Figure 5a.

The daily revenues and profits under original, optimized and benchmark prices are compared in Figure 6. The comparison shows a closely similar revenue in the three cases. The optimized prices have given a slightly smaller revenue compared to the revenue from original and benchmark prices. However, the profit from original prices is considerably smaller than the profit from optimized prices. The latter is only slightly smaller than the benchmark's profit. Numerically, the profit from the proposed methods is 95.97% of the optimal benchmark profit. This observation shows that an increase in the profit can be made without an increase in the revenue when the prices are optimized correctly.

Discussion and conclusion

In this paper, we demonstrated the effectiveness of a new TCL control using electricity price proxies. The control policy

Figure 3. Learning process of a population of size 100.

Figure 4. Results' comparison of original and optimized pricing policy. (a) Optimized prices solution for 24 hours. (b) Revenue and profit under original and optimized prices for 24 hours. (c) Total electricity consumption under original and optimized prices. (d) Daily electricity bills under original and optimized prices.

Figure 5. Results' comparison of optimized and benchmark pricing policy. (a) Comparison between benchmark and optimized prices. (b) Hourly revenues and profits under optimized prices and benchmark prices. (c) Hourly total electricity consumption under optimized prices and benchmark prices. (d) Daily electricity bills under optimized and benchmark prices.

Figure 6. Daily revenues and profits under original, optimized and benchmark prices.

consists of a sequence of prices influencing the electricity consumption from TCLs. The problem was formulated as a Markov decision process with non-Markovian state to handle the sparse observations of the TCL cluster's state. We extend the observable state with sequences of past observations to approximate the transition function using an LSTM architecture. The LSTM network is used to capture the individual behavior of TCLs under price-based DR. The individual models are aggregated to approximate the next state of the cluster. This approximation is used iteratively in a genetic algorithm to evaluate the potential profit from an energy arbitrage operation and find the optimal pricing policy for a given control horizon. The LSTM models are updated every 24 hours to capture the changes in the TCL units' behavior.

The experiment consists of a retailer agent buying electricity from the wholesale market and selling it to a group of residential TCLs. The agent can only measure the electricity consumption of each TCL and the outside temperature. The agent has access to a significant amount of historical data from an already implemented DR program. Which allows it to train the LSTM models for each TCL unit and perform an optimization on the electricity prices.

We first evaluate the performance of the LSTM network by comparing the real and predicted loads from 30 TCL units during different days. The predicted load profiles are closely similar to the real load profiles both at individual and aggregate level. The optimization relies on a genetic algorithm with a profit maximization objective. The results of the optimization show that the proposed methods offer a much higher daily profit than the original prices and 95.97% of the optimal profit from a model that has full observation of the state.

The flexibility offered by TCLs is a high potential for ancillary services required for a deep integration of renewable

energy sources in the grid. An energy arbitrage operation can offer a service to the grid by exploiting this flexibility using direct or indirect control. The partially observable state and the uncertainty of the TCL response to prices was tackled in this paper with an LSTM network using past observations and actions. The LSTM network offered a high performance by extracting relevant features of the hidden state using its internal memory cell, allowing it to process sequences of sparse observations to learn the hidden patterns of power consumption.

Data availability

Underlying data

Figshare: LSTM+GA data, <https://doi.org/10.6084/m9.figshare.9746786.v1>⁴⁴.

This project contains the following underlying data:

- Data used by the fuzzy logic simulation model such as temp_prices and temperatures.
- Data generated by the fuzzy simulator such as fuzzy_outxx.csv and used to train the LSTM models.
- Data related to the optimization process such as results and GA_pricing, optimized_prices_loads

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Software availability

Code for analysis available from: <https://github.com/tahanakabi/Optimal-Price-Based-control-of-heterogeneous-thermostatically-controlled-loads-under-uncertainty-usi>

Archived code as at time of publication: <http://doi.org/10.5281/zenodo.3383615>⁴⁵

License: MIT

References

1. Halamayi DA, Brekken TKA, Simmons A, *et al.*: **Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation.** 2010. [Publisher Full Text](#)
2. Mathieu JL, Kamgarpour M, Lygeros J, *et al.*: **Arbitraging Intraday Wholesale Energy Market Prices With Aggregations of Thermostatic Loads.** *IEEE Transactions on Power Systems.* 2015; **30**(2): 763–772. [Publisher Full Text](#)
3. D&R International, Ltd.: **2011 Buildings Energy Data Book.** 2012. [Reference Source](#)
4. U. E. I. Administration: **U.S. Energy Information Administration.** 2010. [Reference Source](#)
5. Mathieu JL, Kamgarpour M, Lygeros J, *et al.*: **Energy arbitrage with thermostatically controlled loads.** 2013. [Accessed 11 6 2019]. [Publisher Full Text](#)
6. Maasoumy M, Razmara M, Shahbakhti M, *et al.*: **Selecting building predictive control based on model uncertainty.** 2014. [Accessed 11 6 2019]. [Publisher Full Text](#)
7. Koch S, Mathieu JL, Callaway DS: **Modeling and Control of Aggregated Heterogeneous Thermostatically Controlled Loads for Ancillary Services.** 2011. [Accessed 11 6 2019]. [Reference Source](#)
8. Saha A, Kuzlu M, Pipattanasomporn M, *et al.*: **Enabling Residential Demand Response Applications with a ZigBee-Based Load Controller System.** 2016; **2**(4): 303–318. [Accessed 11 6 2019]. [Publisher Full Text](#)
9. Verbon GPJ, Beemsterboer S, Sengers F: **Smart grids or smart users? Involving users in developing a low carbon electricity economy.** *Energy Policy.* 2013; **52**: 117–125. [Publisher Full Text](#)
10. Yan X, Ozturk Y, Hu Z, *et al.*: **A review on price-driven residential demand response.** *Renew Sust Energ Rev.* 2018; **96**: 411–419. [Publisher Full Text](#)
11. Hansen M, Borup M: **Smart grids and households: how are household consumers represented in experimental projects?** *Tech Anal Strat Manag.* 2018; **30**(3): 255–267. [Publisher Full Text](#)
12. Littman ML: **Markov Decision Processes.** 2001. 9240–9242. [Accessed 12 6 2019]. [Publisher Full Text](#)

13. Nakabi TA, Toivanen P: **An ANN-based model for learning individual customer behavior in response to electricity prices.** *Sustainable Energy, Grids and Networks.* 2019; **18**.
[Publisher Full Text](#)
14. Ihara S, Schweppe FC: **Physically based modeling of cold load pickup.** *IEEE Transactions on Power Apparatus and Systems.* 1981; **100**(9): 4142–4150.
[Publisher Full Text](#)
15. Malhame R, Chong CY: **Electric load model synthesis by diffusion approximation of a high-order hybrid state stochastic system.** *IEEE TRANSACTIONS ON AUTOMATIC CONTROL.* 1985; **30**(9): 854–860.
[Publisher Full Text](#)
16. Mathieu JL, Koch S, Callaway DS: **State estimation and control of electric loads to manage real-time energy imbalance.** *IEEE Trans Power Syst.* 2013; **28**(1): 430–440.
[Publisher Full Text](#)
17. Hao H, Sanandaji BM, Poolla K, *et al.*: **Aggregate Flexibility of Thermostatically Controlled Loads.** *IEEE Transactions on Power Systems.* 2015; **30**(1): 189–198.
[Publisher Full Text](#)
18. Kamgarpour M, Ellen C, Soudjani SEZ, *et al.*: **Modeling options for demand side participation of thermostatically controlled loads.** In *2013 IREP Symposium Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid.*, Rethymno, Greece, 2013.
[Publisher Full Text](#)
19. Meng FL, Zeng XJ: **A Profit Maximization Approach to Demand Response Management with Customers Behavior Learning in Smart Grid.** *IEEE Trans Smart Grid.* 2016; **7**(3): 1516–1529.
[Publisher Full Text](#)
20. Jia L, Zhao Q, Tong L: **Retail pricing for stochastic demand with unknown parameters: An online machine learning approach.** In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton).* Monticello, IL USA, 2013.
[Publisher Full Text](#)
21. Dehghanpour K, Nehrir HM, Sheppard JW, *et al.*: **Agent-Based Modeling of Retail Electrical Energy Markets With Demand Response.** *IEEE Transactions on Smart Grid.* 2018; **9**(4): 3465–3475.
[Publisher Full Text](#)
22. Xie D, Hui H, Ding Y, *et al.*: **Operating reserve capacity evaluation of aggregated heterogeneous TCLs with price signals.** *Applied Energy.* 2018; **216**: 338–347.
[Publisher Full Text](#)
23. Jay D, Swarup K: **Price Based Demand Response of Aggregated Thermostatically Controlled Loads For Load Frequency Control.** In *17TH NATIONAL POWER SYSTEMS CONFERENCE.* 2012.
[Reference Source](#)
24. Wang P, Zou S, Wang X, *et al.*: **A Stackelberg Game Approach for Price Response Coordination of Thermostatically Controlled Loads.** *Applied Sciences.* 2018; **8**(8): 1370.
[Publisher Full Text](#)
25. De Paola A, Trovato V, Angeli D, *et al.*: **A Mean Field Game Approach for Distributed Control of Thermostatic Loads Acting in Simultaneous Energy-Frequency Response Markets.** *IEEE Transactions on Smart Grid.* Early Access 2019; 1–1.
[Publisher Full Text](#)
26. Grammatico S, Gentile B, Parise F: **A Mean Field control approach for demand side management of large populations of Thermostatically Controlled Loads.** In *2015 European Control Conference (ECC).* Linz, Austria, 2015.
[Publisher Full Text](#)
27. Nakabi TA, Haataja K, Toivanen P: **Computational Intelligence for Demand Side Management and Demand Response Programs in Smart Grids.** In *8th International conference on bioinspired optimization methods and their applications.* Paris, 2018.
[Reference Source](#)
28. Can Kara E, Berges M, Krogh B, *et al.*: **Using smart devices for system-level management and control in the smart grid: A reinforcement learning framework.** In *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm).* Tainan, Taiwan, 2012.
[Publisher Full Text](#)
29. De Somer O, Soares A, Kuijpers T, *et al.*: **Using Reinforcement Learning for Demand Response of Domestic Hot Water Buffers: a Real-Life Demonstration.** Cornell University, 2017.
[Reference Source](#)
30. Ruelens F, Claessens BJ, Quaiyum S, *et al.*: **Reinforcement Learning Applied to an Electric Water Heater: From Theory to Practice.** *IEEE TRANSACTIONS ON SMART GRID.* 2018; **9**(4): 3792–3800.
[Publisher Full Text](#)
31. Claessens BJ, Vrancx P, Ruelens F: **Convolutional Neural Networks for Automatic State-Time Feature Extraction in Reinforcement Learning Applied to Residential Load Control.** *IEEE Transactions on Smart Grid.* 2018; **9**(4): 3259–3269.
[Publisher Full Text](#)
32. Ruelens F, Claessens BJ, Vandaal S, *et al.*: **Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning.** *IEEE Transactions on Smart Grid.* 2017; **8**(5): 2149–2159.
[Publisher Full Text](#)
33. Ruelens F, Claessens BJ, Vrancx P, *et al.*: **Direct Load Control of Thermostatically Controlled Loads Based on Sparse Observations Using Deep Reinforcement Learning.** Cornell University. 2017.
[Reference Source](#)
34. Patyn C, Ruelens F, Deconinck G: **Comparing neural architectures for demand response through model-free reinforcement learning for heat pump control.** In: *2018 IEEE International Energy Conference (ENERGYCON).* Limassol, Cyprus, 2018.
[Publisher Full Text](#)
35. Mocanu E, Constantin Mocanu D, Nguyen PH: **On-line Building Energy Optimization using Deep Reinforcement Learning.** *IEEE Transactions on Smart Grid.* 2018; **10**(4): 3698–3708.
[Publisher Full Text](#)
36. Mohsenian-Rad AH, Wong VW, Jatskevich J, *et al.*: **Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid.** *IEEE Transactions on Smart Grid.* 2010; **1**(3): 320–331.
[Publisher Full Text](#)
37. Holland JH: **Genetic Algorithms.** *Scientific American.* 1992; **267**(1): 66–72.
[Publisher Full Text](#)
38. Bickel T, Thiele L: **A comparison of selection schemes used in evolutionary algorithms.** *Evol Comput.* 1996; **4**(4): 361–394.
[Publisher Full Text](#)
39. Syswerda G: **Uniform crossover in genetic algorithms.** *Proceedings of the 3rd International Conference on Genetic Algorithms.* San Francisco, CA USA, 1989.
[Reference Source](#)
40. Neubauer A: **Adaptive non-uniform mutation for genetic algorithms.** In: *Computational Intelligence Theory and Applications.* Berlin, Heidelberg, 1997; 24–34.
[Publisher Full Text](#)
41. Deb K: **An efficient constraint handling method for genetic algorithms.** *Comput Methods Appl Mech Eng.* 2000; **186**(2–4): 311–338.
[Publisher Full Text](#)
42. **Weather observations, Kaisaniemi observation station Helsinki.** Finnish meteorological institute. [Accessed 8 September 2019].
[Reference Source](#)
43. **Nord Pool, Elspot Day-ahead, Prices.** [Accessed 8 September 2018].
[Reference Source](#)
44. Nakabi T: **LSTM+GA data.** figshare. Dataset. 2019.
<http://dx.doi.org/10.6084/m9.figshare.9746786.v1>
45. Nakabi TA: **tahanakabi/Deep-Reinforcement-learning-for-TCL-control: First release (Version V1.0.0).** *Zenodo.* 2019.
<http://dx.doi.org/10.5281/zenodo.3383615>

Open Peer Review

Current Peer Review Status: ? ? X

Version 1

Reviewer Report 01 September 2020

<https://doi.org/10.5256/f1000research.22445.r68826>

© 2020 Swarup S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Shanti Swarup

Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

1. The idea behind TCL is very good and how it effects the price.
2. However, the paper is not properly presented.
3. What is LSTM? long-short-term memory is contradicting what is long -short. Either it has to be long term memory (LTM) or short-term Memory (STM).
4. Discussions and conclusions should be separate. Difficult to identify the conclusions. It should be results and discussions.
5. There are three different tools employed as shown below. Why is there a need to use all these tools (DL, LTSM, GA)
Deep learning is used for control of TCL loads
LSTM networks for state estimation
Genetic algorithms for price optimization
Only one tool can be used.
In fact; prediction of load for TCL is missing.
6. Why is a need for price optimization?
7. The price (LMP) is dependent on the intersection between the generation and demand. This price keeps on varying.
8. The social benefit or social welfare needs to be optimized and not the price. Eqn 6 is not correct.
9. Figs 2a and 2b do not provide sufficient information to infer the contribution. The need for

so many plots is questionable.

10. Only important results should be provided.

11. In spite of the good idea and motivation, the approach used seems to be not proper.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

No

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Demand Response and Management

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 01 September 2020

<https://doi.org/10.5256/f1000research.22445.r68823>

© 2020 Huang C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chiou-Jye Huang

Department of Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China

The authors proposed a long-short-term memory (LSTM) network to learn the individual behaviors of TCL units. The authors use the aggregated information to predict the response of the TCL

cluster to the pricing policy. The authors use this prediction model in a genetic algorithm to find the best prices in terms of profit maximization in an energy arbitrage operation. The simulation results show that the proposed method offers a profit equal to 96% of the theoretically optimal solution. I recommend minor revisions. I recommend the following revisions. In addition, there are some questions that need to be explained below:

1. English language should be carefully checked and carefully check paper for language typos.
2. Some figures are not needed.
3. All the figures are unclear and hard to read, please update to a clear version.
4. The authors must provide a detailed flowchart of the methodology of the paper.
5. The conclusion section is missing some perspective related to future research work.
6. References are too few and must be updated in recent years. I suggest authors should add related references.

References

1. Huang C, Kuo P: Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting. *IEEE Access*. 2019; **7**: 74822-74834 [Publisher Full Text](#)
2. Kuo P, Huang C: An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. *Sustainability*. 2018; **10** (4). [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Big data analysis, machine learning and deep learning applications on the Internet of Energy (IoE) and environmental science, especially in renewable energy, as well as

electricity load demand, electricity prices, solar radiance, photovoltaic power, and PM2.5 forecasting, and photovoltaic power plants planning design, and operation maintenance management.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 25 August 2020

<https://doi.org/10.5256/f1000research.22445.r68829>

© 2020 Shun M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Matsukawa Shun

Smart Grid Power Control Engineering Joint Laboratory, Gifu University, Gifu, Japan

1. The sequence length of the LSTM model is so short that it does not show superior predictive ability compared to other models.
2. I couldn't understand the significance of Figure 2 because of its lower resolution. This issue needs to be resolved.
3. On my own interpretation of Figure 2, I felt that the orange breakline showing the LSTM prediction results failed to learn the response of the TCL agent because it was heavily dependent on the load of the previous step.
4. In Figure 4 and 5, the optimized loads on time 14-15 are changes suddenly. I felt it is necessary to investigate the cause of it.

Comment) I felt that the real key to this study was not the optimization, but the accuracy of TCL response prediction. Therefore, a more detailed analysis of the LSTM model would further enhance the value of this paper.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: smart grid, baseline load estimation, machine learning, neural networks, time-series, LSTM

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research