LSTM-Embedded Q-learning for Daily Operation of Microgrid with Renewable Uncertainty and Demand-Side Management

Zeina Bahij, Bing Yan

***Abstract*— The daily operation of microgrids is essential to reduce costs and avoid possible severe damage to the system. The problem is challenging, especially at the sub-hourly level, due to the intermittency of variable renewable generation, the uncertainty of demand, and the tight coupling between the supply and demand sides. In this paper, a mixed-integer linear programming model is established for microgrid daily operation considering demand-side management through load shifting. Moreover, a novel Long Short-Term Memory embedded Q-learning (LSTM-Embedded Q-learning) model is developed to predict the operation strategy (energy device scheduling and load shifting). To address the issue of intermittency of variable renewable generation and uncertainty of demand, LSTM is embedded into the deep Q-learning (replacing the deep neural network) while utilizing long-term dependencies within sequential input data. In addition, to handle the tight coupling between supply and demand, domain knowledge is incorporated into Q-learning to avoid infeasibility by masking invalid operation strategies and assigning a large penalty. The model has been tested on a microgrid consisting of combined cooling, heating and power, heat pump, photovoltaic, battery, and thermal storage systems. Results show that our model outperforms LSTM (Long-Short Term Memory), state-action-state-reward (SARSA), and Deep Q-learning (DQN) on finding optimized operation strategies during different days of the year, especially on a day with highly variable weather conditions.**

***Index Terms*—Microgrid operation, Renewables, Deep reinforcement learning, LSTM, Q-learning**

I. INTRODUCTION

Microgrids are small-scale power grids that generate electric and thermal energy for a localized area. Establishing the day-ahead operation strategy is a critical problem; it involves generation devices and storage units scheduling, and load shifting to achieve minimum cost and avoid possible severe damage to the system. The problem is challenging due to the intermittency of variable renewable generation and the uncertainty of demand, especially during days with highly variable weather conditions. Also, the tight coupling between the supply and demand incorporates interdependencies, e.g., the devices are scheduled based on the shifted demand.

Managing the operation of microgrids has been studied by many researchers. At the state of practice, mathematical methods find the operation strategy, where renewable generation is modeled by either the deterministic approach or the stochastic approach. In the deterministic approach, renewable generation is derived from physical models without considering uncertainties [1]. In the stochastic approach, renewable generation and demand are modeled by their probability distribution functions without considering their change with the time of the day and the season of the year [2]. Mathematical methods rely on explicit scenario generation or probabilistic modeling as inputs to solve for the operation strategy [3]. In addition, for real-time daily operation with finer time steps, the optimization problems must be repeatedly re-solved with updated forecasts, possibly leading to long solving times. For machine learning models, they can directly predict the operation strategy by learning from the input data of historical renewable generation and demand to capture their random fluctuations. Once trained, they can predict the operation strategy fast and promptly respond to the updated input data [4]. During days with highly variable weather conditions, the high forecasting error in the input data leads to improper scheduling of devices. When considering the demand-side management (DSM) along with the supply-side management, some papers solved the problems iteratively, as done in [5-6], where generation devices are scheduled first and then the load is shifted to off-peak hours to minimize the cost. In contrast to iterative methods, a simultaneous optimization approach solves the generation unit scheduling and load shifting problems within a single, integrated model [7,8]. Sian et al. [9] reviewed DSM integration with device scheduling simultaneously, emphasizing that optimal DSM coordination enhances total cost minimization.

In this paper, a mixed-integer linear programming (MILP) model is established to minimize operation costs through scheduling energy devices and shifting the shiftable load as presented in Section II. To predict the daily operation strategy of a microgrid while incorporating DSM, a novel Long Short-Term Memory embedded Q-learning (LSTM-Embedded Q-learning) model is developed in Section IV. In order to address the intermittency of the renewable generation and demand during days with highly variable weather conditions, Long Short Term Memory (LSTM) is embedded into the deep Q-learning (DQN), replacing the deep neural network (DNN) to predict the operation strategy while utilizing long-term dependencies within sequential data. Furthermore, to address the tight coupling between the supply and demand, the domain knowledge is incorporated into the LSTM-Embedded Q-learning. The model is restricted to select valid operation strategies that satisfy constraints (e.g., capacity limits of energy devices) by masking invalid generation levels of devices and charge/discharge stored energy levels and assigning a large penalty to these actions.

The model has been tested on a microgrid consisting of combined cooling, heating, and power (CCHP), heat pump (HP), photovoltaic (PV), battery, and thermal storage systems in Section V. The power, cooling, and heating demand data are collected from a residential home in New York for the year of 2020 at a 15-minute timestep [10]. The PV power data is collected from a 2-MW PV farm at Rochester Institute of Technology (RIT) in Rochester, New York, for the same year at 5-minute timestep [11]. Using the actual PV and demand data,

Zeina Bahij and Bing Yan are with the Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA (e-mail: zb6781@rit.edu, bxyeee@rit.edu).

This work is based upon work supported by the National Science Foundation (NSF) under Award ESSC-2340095. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

979-8-3315-8138-1

the operation problem of the microgrid under study is solved by MATLAB Optimization Toolbox (branch-and-cut) to generate the training and testing datasets. The datasets are identified as the energy consumption and generation levels of microgrid devices, the state of storage, and on/off statuses of the uninterruptible shiftable loads, and the to/from shifting timesteps of the shiftable loads for the testing day. Results show that LSTM-embedded Q-learning outperforms LSTM, state-action-state-reward (SARSA), and DQN on finding optimized operation strategies on different days of the year, especially on a highly variable weather day.

II. Problem Formulation

This section gives a summary of the microgrid structure under study and the problem formulation. Subsection A shows the microgrid structure; supply-side management formulation is presented in Subsection B; demand-side management formulation is introduced in Subsection C; power and thermal balances are discussed in Subsection D; and the objective function is presented in Subsection E.

A. Microgrid structure

The microgrid under study consists of a CCHP, HP, PV, battery, and heating and cooling energy storage systems, as shown in Figure 1. Power demand and HP required power are satisfied by CCHP generated power from the gas turbine (GT), PV power, the grid power, and the power from the battery. Space heating (SH) demand is satisfied by the thermal energy provided by the CCHP through the heat recovery boiler (HRB), HP, and the SH storage. Space cooling (SC) demand is satisfied by the thermal energy provided by the CCHP through the absorption chiller (abs), HP, and the SC storage.

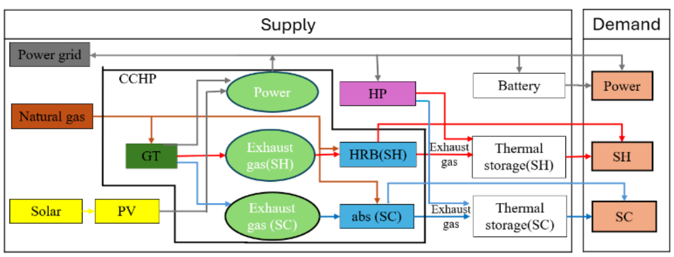

Figure 1: Microgrid structure

B. Supply-side management

Devices are scheduled by determining the ON/OFF statuses of generation devices and their generation levels, the charging/discharging statuses of storage and their corresponding charging/discharging rates. For all the devices except for PV, battery, and thermal storage, there is a binary variable x_{dev} to present the device's on/off status. If the device is on ($x_{dev}(t)$ is 1), its generation level has to be within the minimum and maximum levels. Constraints of devices are presented below:

1. CCHP: CCHP consists of a GT, an HRB, and an abs. The volumetric flow rate of natural gas in the GT is presented in (1),

$$G_{ICE}(t) = P_{CCHP}/\eta_e LHV_{gas}, \tag{1}$$

where P_{CCHP} is the power provided by the GT, η_e is the engine gas-to-electric efficiency, and LHV_{gas} is the lower heat value of natural gas.

In addition, between two consecutive timesteps, the change in power level should not exceed ramp-up or go below ramp-down rates. The heat rate available from the exhaust gas recovered from the gas turbine is shown in Eq. (2). This exhaust gas is recovered by the HRB and the abs to generate SH and SC as in Eq. (3) and Eq. (4), respectively.

$$Q_{GT,ex}(t) = P_{GT}(t)(1-\eta_e-\mu_{GT})/\eta_e, \tag{2}$$

$$H_{HRB,ex}(t) = Q_{GT,ex}(t).\ \xi_{SH}(t).\eta_{HR,HRB}\ .\ COP_{HRB}, \tag{3}$$

$$C_{abs,ex}(t) = Q_{GT,ex}(t).\ \xi_{SC}(t).\eta_{HR,abs}\ .\ COP_{abs}, \tag{4}$$

where μ_{GT} is the present heat loss of the GT, $\xi_{SH}(t)$and $\xi_{SC}(t)$ are the fractions of exhaust gas for SH and SC; their summation should be 1, $\eta_{HR,HRB}$ and $\eta_{HR,abs}$ are the waste heat recovery efficiency, and COP_{HRB} and COP_{abs} are the coefficients of performance of the HRB and the abs.

The volumetric flow rate of natural gas by the HRB:

$$G_{HRB,di}(t) = H_{HRB,di}(t)/(LHV_{gas}.\eta_{HP}.COP_{HRB}), \tag{5}$$

where $H_{HRB,di}$ is the heating rate generated by natural gas, and η_{HRB} is the efficiency of the HRB.

The flow rate of natural gas is modeled same way for the abs.

2. HP: the power consumption of the HP for heating:

$$P_{HP,H}(t) = (H_{HP}(t)\times\Delta t)/COP_{HP,H}, \tag{6}$$

where $COP_{HP,H}$ is the coefficient of performance of the HP in the heating mode, and $H_{HP}(t)$ is the heating rate in the HP. The cooling in the HP is modeled the same way as in heating.

3. Battery: The state-of-charge (SOC) dynamics:

$$SOC_B(t+1) = SOC_B(t)(1-\delta_B) - \frac{P_{dis}(t)}{\eta_{dis}}\Delta t + \eta_{ch}P_{ch}(t)\Delta t, \tag{7}$$

where $SOC_B(t)$ and $SOC_B(t+1)$ are the SOC of the battery at t and t+1, η_{dis} and η_{ch} are the discharge and charge efficiencies, and δ_B is the energy loss ratio. The battery cannot be charged and discharged simultaneously, and the SOC level should be within the minimum and maximum limits.

4. Thermal Storage: The energy stored for SH:

$$H_s(t) = H_s(t-\Delta t).\left(1-\varphi_s(\Delta t)\right) + \left(H_s^i(t) - H_s^o(t)\right)\Delta t, \tag{8}$$

where φ_s is the heat loss through the tank wall, $H_s^i(t)$ and $H_s^o(t)$ are the heat rates brought in and taken out by the flow in and out of water, and $H_s(t)$ and $H_s(t-\Delta t)$ are the energy stored at the current and previous time steps.

The stored energy for SH should be within the minimum and maximum limits. SC storage is modeled the same way as the SH storage.

C. Demand-Side Management

Since the supply-side is managed simultaneously with the demand, shifting decisions will be also based on generation availability and grid power price. The interdependency between the two problems is presented in the power balance constraint (to be shown later), where generation is equal to the optimized total demand. In this study, two kinds of shiftable loads are considered: the shiftable interruptible load that is presented by $P^s(t)$ and can operate with interruptions and the uninterruptible shiftable load that is presented by $P_a^u(t)$ for each appliance *a* and should operate for a specific duration once on. The constraints of the DSM are presented as below:

1.Interruptible shiftable loads: The interruptible shiftable load is shifted from original t_{fr} to t_{t0} with interruptions. Interruptible shiftable load constraints are presented below:

a) Non-simultaneous load addition and removal: To avoid simultaneous addition and removal of the shiftable load at the same hour, binary decision variables $u_x[t_{t0}]$ to present if the load is removed from t_{t0} and $v_x[t_{t0}]$ to present if the load is added to t_{t0} have been assigned. The nonsimultaneous addition and removal of load to t_{t0} is presented in Eq. (9),

$$u_x[t_{t0}] + v_x[t_{t0}] \le 1, \forall t_{t0}. \tag{9}$$

b) Peak load constraint: The post-optimization load is the sum of the net shifted load, the non-shiftable load, and the shiftable load that is not shifted. This load cannot exceed the original peak load P^P at a timeslot as presented in Eq. (10),

979-8-3315-8138-1

$\sum_{t_{fr}}^{T} x\left[t_{fr}\right][t_{t0}] - \sum_{t_{fr}}^{T} x\left[t_{fr}\right][t_{t0}] + P^{s}(t_{t0}) + P^{ns}(t_{t0}) + P_{a}^{u}(t) \leq P^{P}, \forall\, t_{t0}$. (10)

c) Minimum and maximum shiftable load constraints: The load being shifted from time t_{fr} to t_{t0} must be within the range of zero to the maximum available shiftable load present at the t_{fr} ($u\left[t_{fr}\right][t_{t0}]P^{s}(t_{t0})$). Also, the load that is being shifted from t_{t0} to t_{fr} should be greater than zero and less than the available shiftable load at the t_{t0} ($v[t_{t0}]\left[t_{fr}\right] P^{s}(t_{t0})$).

d) Total shifting constraint: The sum of load shifted from an t_{fr} cannot exceed the total shiftable load at that t_{fr}:

$\sum_{t}^{T} x\left[t_{fr}\right][t_{t0}] \leq P^{s}(t_{fr}), \forall t_{fr}$. (11)

e) Self-shifting constraint: The shiftable load at t_{t0} cannot be added to its own t_{t0}:

$x[t_{t0}][t_{t0}] = 0, \forall t_{t0}$. (12)

2.Uninterruptible shiftable loads: The uninterruptible shiftable appliances cannot be interrupted once on and should run for a period of d_a. For each uninterruptible shiftable appliance *a,* a binary decision variable $x_a(t)$ to present the on/off status and a start indicator $z_a(t)$ are assigned. The interruptible shiftable appliances constraints are as follows:

a) Start signals when switching on: when $z_a(t)$ is 1, the appliance is on ($x_a(t) = 1$) as presented in Eqn. (13),

$1 - x_a(t-1) \leq z_a(t) \geq x_a(t) \;\; \forall t, a$. (13)

b) Time-window scheduling constraint: the operation of appliances *a* is restricted to specific hours during daytime. $x_a(t) = 0 \;\forall a,\; t \notin T_{day}$,

where T_{day} is a set of daytime time slots (e.g.,8:00- 20:00).

c) Uninterruptible operation: once appliance *a* is on it should run for a period of d_a,

$\sum_{k=t}^{t+d_a-1} x_a(k) \geq d_a z_a(t), \forall a$. (14)

D. Power and thermal balances

A power balance among $P_d^{opt}(t)$ that is optimized demand after optimization, generation, battery charge/discharge, and grid power is established as shown in Eq. (15),

$P_d^{opt}(t) + P_{ch}(t) + P_{HP}(t) + P_{CP}(t) + P_s(t) = P_{dis}(t) + P_{PV}(t) + P_{GT}(t) + P_b(t)$, (15)

where $P_s(t)$ and $P_b(t)$ are the sold and bought grid power $P_d^{opt}(t)$ is the summation of shiftable and non-shiftable load after optimization as shown in Eqn. (16),

$P_d^{opt}(t) = \sum_{t_{fr}}^{T}(x\left[t_{fr}\right][t_{t0}] - x[t_{t0}]\left[t_{fr}\right]) + P^{s}(t_{t0}) + P^{ns}(t_{t0})$. (16)

Heating demand $H_d(t)$ and cooling demand $C_d(t)$ should be satisfied by the energy devices generation, as shown in Eq. (17) and (18),

$H_d(t) = H_{CCHP}(t) + H_{HP}(t) + H_s^{o}(t) - H_s^{i}(t)$, (17)

$C_d(t) = C_{CCHP}(t) + C_{CP}(t) + C_s^{o}(t) - C_s^{i}(t)$. (18)

E. Objective Function

This work aims at minimizing the operation cost, which is the grid power cost, the difference between buying power from the grid and selling power to the grid, and the natural gas cost as in Eq. (19),

$C_T = \sum_{t=1}^{T}\left(c_b(t)P_b(t)\Delta t - c_s(t)P_s(t)\Delta t + c_g G_{CCHP}(t)\right)$, (19)

where c_b and c_s are the cost/revenue of buying/selling power from/to the grid, and c_g is the cost of gas.

III. Methodology

The LSTM-Embedded Q-learning model is developed in this section. To address the intermittency of the renewable generation and demand, LSTM is embedded into the deep Q-learning (replacing the DNN) to predict the operation strategy while utilizing long-term dependencies within sequential data as presented in Subsection A. In addition, to handle the tight coupling between the supply and demand, the domain knowledge is incorporated into the LSTM-Embedded Q-learning as illustrated in Subsection B. The overall approach is summarized in Subsection C.

A. LSTM-Embedded Q-learning Structure

DQN is a reinforcement learning model that combines Q-learning and DNN, allowing an agent to learn the optimal policy (mapping states to actions) in an environment to achieve an optimal reward. For each possible action *A* in state *S*, the DQN model calculates a *Q-value Q*(*S,A*)*,* which represents the expected future reward the agent can get by taking action *A* in state *S*. The DNN takes states *S* as input and outputs the potential *Q-values* for possible actions *A*. In this study, the DNN is replaced by LSTM to predict the *Q-values* based on sequences of environment and state data rather than a single observation. The structures of the Q-learning and LSTM are presented below.

1. Q-learning Structure: the environment, agents, actions, and reward of the Q-learning are presented as follows*:*

a) Environment: the environmental observation space $E(t)$ is the input data at time *t*, as shown in Eq. (20),

$E(t)\!: \{P_{PV}(t), P^{ns}(t), P^{s}(t), P_a^{u}(t), C_d(t), H_d(t)\}$. (20)

b) Agents: agents *G* are the energy devices, the battery, the thermal storage, interruptible shiftable load and the uninterruptible shiftable appliances *a,* as shown in Eq. (21),

$$G\!:\left\{\begin{matrix} GT, HRB, abs, battery, HP, thermal\ storage, \\ grid\ power, shiftable\ demand, appliances\ a \end{matrix}\right\}. \quad (21)$$

c) Actions: actions $A(t)$ of the supply-side are the generation levels and the on/off statuses of the devices and the charge/discharge power to the storage. For the demand side, actions are the shifting decisions of interruptible shiftable loads and the on/off status of the uninterruptible appliances *a* as in Eq. (22),

$$A(t)\!:\left\{\begin{matrix} P_{GT}(t),\ H_{HRB,di}(t), C_{abs,di}(t), \xi_{SH}(t)\xi_{SC}(t), \\ x_{GT}(t), x_{HRB}(t), x_{abs}(t), H_{HP}(t), C_{HP}(t), \\ x_{HP}(t), P_c(t), P_d(t), x_c(t), x_d(t), H_s^{i}(t), H_s^{o}(t), \\ C_s^{i}(t), C_s^{o}(t), u^{i}\left[t_{fr}\right][t_{to}], v^{i}\left[t_{fr}\right][t_{to}], x_a(t) \end{matrix}\right\}. \quad (22)$$

d) States: for the supply-side, states $S(t)$ are the energy consumption of energy devices and the state of storage. For the demand-side, states are the power shifted at time t_{to} from t_{fr} of the interruptible shiftable loads and the start indicator of the uninterruptible shiftable appliances *a*, as presented in Eq. (23),

$$S(t)\!:\left\{\begin{matrix} G_{GT}(t), G_{HRB,di}(t), G_{abs,di}(t), H_{HRB,ex}(t), \\ C_{abs,ex}(t), P_{HP,H}(t), P_{HP,C}(t), SOC_B(t), \\ H_s(t), C_s(t), P^{s}\left[t_{fr}\right][t_{to}], z_a(t) \end{matrix}\right\}. \quad (23)$$

e) Rewards: the reward is taken to be the negative value of the total daily operation cost as in Eq. (24),

$R1_T = -C_T$. (24)

2. LSTM Structure: LSTM captures long-term dependencies in the sequential data of the historical three days of environmental data *E*(*t*) and states *S*(*t*), and the next day's forecasted data to predict the *Q-values* of the deep Q-learning. The input sequences of the LSTM layer are 6 sequences of actual PV and demand data (presented in equation (20)) along with 12 sequences of states (presented in equation (23)) for the previous three days, starting from timestep -3*T*+1 till 0, and the forecasted data for the next day (i.e., day *d*) starting from

979-8-3315-8138-1

timestep 1 till timestep t. The output will be the *Q-values* for each action.

C. Domain knowledge

When the agent receives a new state, the LSTM processes the sequential states and environmental input and estimates *Q-values* for all possible actions. The action of the highest Q-value should be selected to maximize the final reward. Incorporating domain knowledge involves information about the device's constraints. Rather than choosing actions of the highest *Q-values* that might lead to infeasibility, incorporating domain knowledge helps in better training and more accurate prediction results. Constrained action selection via masking is a technique used in Q-learning to ensure that the agent's chosen actions adhere to domain-specific rules and constraints at every decision. Before selecting the best action, actions that do not match the physical constraints of devices (like minimum and maximum generation limits) will be masked, where their *Q-values* are assigned a penalty (e.g., negative infinity). With large negative *Q-values*, their corresponding actions will not be selected. This forces the agent to consider only feasible actions that comply with operational constraints.

C. Overall Approach

The overall LSTM-Embedded Q-learning approach is depicted in Figure 2. As shown in the Figure, the LSTM is embedded inside the deep Q-learning to predict the *Q-values*, replacing the DNN to process sequences of inputs. At the beginning of the testing day, the LSTM processes sequences of environmental data and states of size (12×3T) starting from the previous three days till the start of the testing day (-3T to 1), and outputs *Q-values* at time step t for each possible action $A(t)$ at given state $S(t)$. In order to handle the tight coupling between supply and demand, the domain knowledge is incorporated by limiting action selection to physically valid ones. The *Q-values* for invalid actions are masked by assigning them a very low value. The optimal actions of size 22×1 are then chosen based on the highest *Q-value*, and the environment and states data are shifted by one time step t (-3T+t to t) and passed again to the LSTM as inputs. The process is repeated till the end of the testing day when t=T.

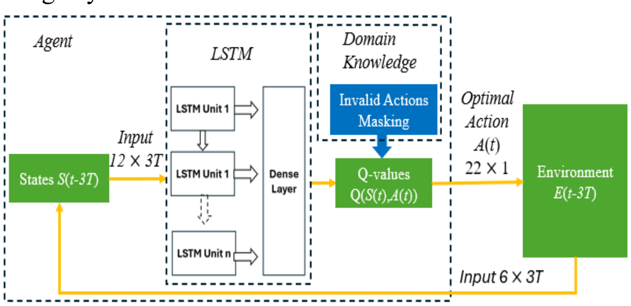

Figure 2: LSTM-Embedded Q-learning

The prediction error is calculated based on the total operation cost. After predicting the operation strategy for the whole day and specifying the energy consumption of each device, the operation cost is obtained using Eq. (19). To test the model performance, the correlation (r), mean absolute error (*MAE*), and root mean square error (*RMSE*) are calculated between the predicted cost via LSTM-embedded Q-learning and the values obtained by mathematical optimization.

IV. Results

The LSTM-Embedded Q-learning model is tested in this section. The microgrid's configuration is introduced in Subsection A, and the overall performance and comparison with other models are presented in Subsection B.

A. Microgrid Configuration

1. The Microgrid Configuration and Devices Size: The sizes of devices, except for the PV, are determined based on the total peak demand to be presented later. GT generation limits are 10-50 kW, battery charging/discharging power limits are 5-20 kW, HP heating/cooling limits are 5-20 kWh, and the abs and HRB limits are 5-30 kWh.

2. Demand: Power, heating, and cooling demand data are taken from a large residential home in New York State at a 15-minute timestep for the year 2020 [10].

3. PV: The PV power data is taken from a 2-MW solar farm from Rochester Institute of Technology in New York for the year 2020 [11]. The data is at a 5-minute timestep of a total of 105,120 data points. The power generation is scaled down 10 times to adjust to the power demand of the residential home under study. To match the demand data, the PV data points at every 15 minutes are systematically extracted every third data point from the original 5-minute data.

4. Grid Power and Natural Gas Prices: The price of natural gas c_g is 0.180 \$/kWh for the whole day. The cost of buying power from the grid c_b is 0.21 \$/kWh during peak hours 9:00 am to 6:00 pm [12] and 0.14\$/kWh otherwise. The price of selling power to the grid c_s is set as 80% of the buying cost c_b.

5. Training and Testing Datasets: With the collected demand, PV power, and the grid power and natural gas prices, the operation optimization of the microgrid under study is solved using the branch-and-cut method implemented in MATLAB 2025 Optimization Toolbox due to its effectiveness in solving MILP problems. The problem has been solved each day separately for the year 2020, and the cost is obtained. The dataset is divided as 80% for training, 10% for validation, and 10% for testing from each month to consider all possible changes in the weather conditions.

B. LSTM-Embedded Q-Learning Testing Results

1. Overall Performance: After training the LSTM-Embedded Q-learning model, it shows a good accuracy in predicting the daily operation strategy. The prediction time is 10.34 seconds. The correlation between the cost obtained by math optimization using actual PV and demand data and the cost predicted by the LSTM-Embedded Q-learning is 96.86% using the training dataset and 95.03% using the testing dataset. This demonstrates the effectiveness of the model in predicting the operation strategy while addressing the intermittency of PV and demand and the tight coupling of the supply-demand optimization.

The generation and storage levels of devices and optimized total power demand have been plotted in Figure 3 for a day with highly variable weather conditions (a day with high variance in PV power) to test the model behavior under rapid fluctuations in PV and demand. As shown in Figure 3, during days with highly variable weather conditions, PV power shows rapid fluctuations, and the power heating and cooling demands are variable throughout the day. When applying the LSTM-Embedded Q-learning, the grid power is used when low purchase cost is low (00:00 to 9:00 am and 6:00 pm to 00:00 am), and the CCHP is used for generating thermal energy instead of the HP when the grid price is expensive.

To verify the effectiveness of incorporating the demand-side management into the operation, the operation cost with and without demand-side management has been plotted in Figure 4.

979-8-3315-8138-1

As shown in the figure, the demand is shifted to times of lower grid price, which reduces the daily operation cost.

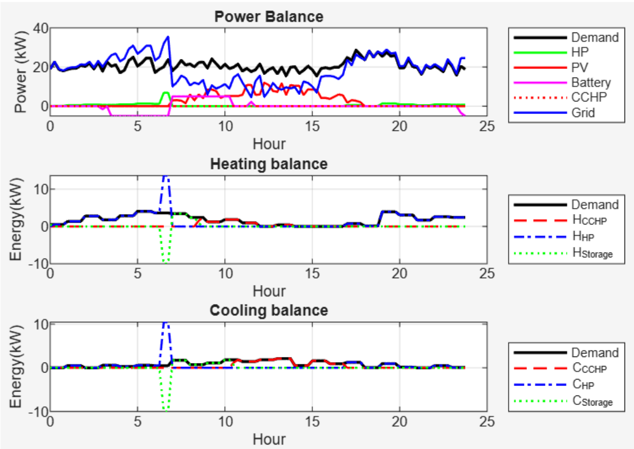

Figure 3: Power and thermal balance

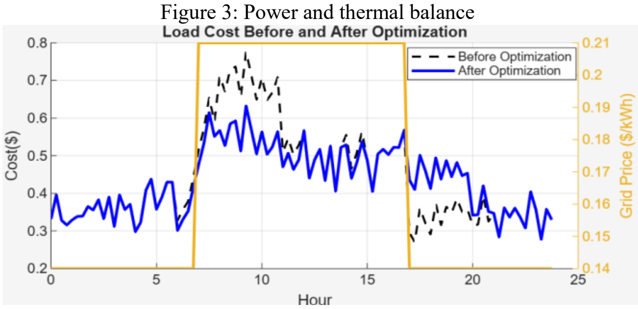

Figure 4: Load cost before and after optimization

2. Comparison with other models: the LSTM-Embedded Q-learning performance has been compared with LSTM, SARSA, and DQN using the testing dataset, as shown in Table 1. As shown in Table 1, LSTM-Embedded Q-learning outperforms other models with a correlation between predicted and actual cost of 95.03%, compared to 83.13% and 87.3% and 90.24% using LSTM, SARSA, and DQN, respectively.

Table 1: Comparison using the testing dataset

	R (%)	**MAE (₵)**	**RMSE(₵)**
LSTM	83.13	244.51	392.21
SARSA	87.30	192.32	325.64
DQN	90.24	145.64	280.15
Ours	**95.03**	**92.18**	**164.92**

To evaluate the model performance in different weather conditions, the prediction accuracy has been tested on a summer day, a winter day, and a day with highly variable weather conditions, as shown in Table 2. LSTM-Embedded Q-learning outperforms all other models on all testing days. The most significant accuracy improvement is recorded on a day with highly variable weather conditions. The predicted correlation using LSTM-Embedded Q-learning is 92.04%, compared to 78.10%, 82.41%, and 87.35% using LSTM, SARSA, and DQN, respectively. During days with highly variable weather conditions, the PV power and the thermal demand of both cooling and heating are intermittent, making the problem even more difficult compared to other days. This leads to high forecasting errors using time series models since the sequence data is not uniform. Also, SARSA and DQN models cannot make accurate decisions based on highly variable data. The LSTM-Embedded Q-learning enhances the prediction accuracy of solutions under historical dependencies, sequential decision-making and tight coupling.

Table 2: Comparison of summer, winter, and variable weather

Day	Summer			Winter			Variable Weather		
	R	MAE	RMSE	R	MAE	RMSE	R	MAE	RMSE
LSTM	87.74	153.45	314.62	89.52	428.01	634.71	78.10	532.44	745.62
SARSA	93.25	114.67	235.90	94.24	336.82	520.99	82.41	345.62	533.47
DQN	95.04	94.63	183.44	95.98	212.44	384.26	87.35	284.56	403.12
Ours	**97.58**	**77.25**	**134.11**	**98.54**	**141.88**	**273.34**	**92.04**	**112.88**	**215.41**

V. Conclusions

The microgrid operation optimization problem is challenging due to the uncertain nature of PV and demand, especially during days with highly variable weather conditions. To address this issue, a novel LSTM-Embedded Q-learning model is developed to predict the operation strategy for the next 24 hours with 15- minutes as time steps. The key idea is to add the last three days of actual input data to the model since consecutive days' weather conditions are close to each other. Also, to address the tight coupling between the supply and demand sides, the domain knowledge is incorporated into the Q-learning to avoid infeasibility. To restrict action selection to physically valid ones that satisfy physical constraints of devices, the idea is to mask invalid actions by assigning large penalty. The model has been tested on a microgrid consisting of CCHP, HP, PV, battery, and thermal storage. Results show that our model outperforms LSTM, SARSA, and DQN by predicting operation strategies close to the optimal solutions obtained by mathematical optimization using the actual PV and demand on testing days from different months. The future work will focus on the multi-objective operation of microgrids to minimize operation cost and emissions. For better prediction, a moving window training method will also be explored, where the actual data, once available, will be incorporated in the sequential input data of LSTM-Embedded Q-learning.

References:

[1] H. Li, Z. Xu, and J. Zhao, “Optimal operation of microgrids with renewable generation: A deterministic approach,” *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6581–6592, 2018.
[2] M. Mohammadi, A. Soleymani, and M. Rastegar, “Stochastic optimization of microgrid operation considering renewable uncertainties,” *Energy*, vol. 118, pp. 827–839, 2017.
[3] A. Recalde, R. Cajo, W. Velasquez, and M. S. Alvarez Alvarado, “Machine Learning and Optimization in Energy Management Systems for Plug In Hybrid Electric Vehicles: A Comprehensive Review,” Energies, vol. 17, no. 13, Jun. 2024.
[4] R. R. Hossain and R. Kumar, “Machine Learning Accelerated Real Time Model Predictive Control for Power Systems,” IEEE/CAA Journal of Automatic Sinica, vol. 10, no. 4, pp. 916–930, Apr. 2023.
[5] A. Parisio and L. Glielmo, “A mixed integer linear formulation for microgrid operation optimization,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 950–957, 2013.
[6] J. Zhao, F. Wen, and Z. Y. Dong, “Flexible demand-side management for microgrid operation optimization,” *IET Gener. Transm. Distrib.*, vol. 9, no. 8, pp. 766–773, 2015.
[7] M. Shafie-Khah and J. P. S. Catalão, “A stochastic multi-layer agent-based model to study the impact of demand response programs in electricity markets,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1136–1146, 2013.
[8] A. Khodayar, M. Shahidehpour, and L. Wu, “Enhancing microgrid operation through simultaneous scheduling of electric and thermal resources,” *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1101–1110, 2013.
[9] S. Sian, R. Yan, and T. K. Saha, “Integration of demand-side management and generation scheduling in microgrids: A review,” *Renew. Sustain. Energy Rev.*, vol. 156, p. 112013, 2022.
[10] Department of Energy, National Renewable Energy Laboratory, “Updated U.S. Low-Temperature Heating and Cooling Demand by County and Sector,” July 2023.
[11] RIT’s New Solar Energy Farm, “2-megawatt array among largest for any NY college” Aug. 14, 2015.
[12] Nyserda, energy prices “New York 2020 Residential process”, Jan.2020.