Learn Dynamic Hosting Capacity Based on Voltage Sensitivity Analysis

Jiaqi Wu, Jingyi Yuan, Yang Weng, and Raja Ayyanar
School of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, USA
{jiaqiwu1, jyuan46, yang.weng, rayyanar}@asu.edu

2023 IEEE Power & Energy Society General Meeting (PESGM) | 978-1-6654-6441-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PESGM52003.2023.10252543

***Abstract*—The extensive use of distributed energy sources (DERs) presents the substantial design, planning, and operational issues for distribution systems, thus prompting the broad adaption of methodologies for photovoltaics (PV) hosting capacity analysis (HCA). Traditional HCA methods require running power flow analysis iteratively, typically in the time-series scenario, to consider the dynamic pattern. However, the time-consuming HCA techniques fail to offer online prediction in large distribution networks because of the computational burden. To tackle the computation challenge, we first provide a deep learning-based problem formulation for HCA, which performs offline training and calculates hosting capacity in real time. The applicable learning model, long short-term memory (LSTM), uses historical time-series data to identify the underlying periodic patterns in distribution systems. However, the accuracy of HC estimation is low in the LSTM without considering system spatial information correlated with HC. To capture such spatial correlation from system measurements, we design dual forget gates in the LSTM and propose a novel Spatial-Temporal LSTM. Moreover, as voltage violations are observed to be one of the most critical constraints of HCA, we construct a voltage sensitivity gate to increase the weight on voltage variation and reduce the mismatch in HC determination. The simulation results on different feeders, such as IEEE 123-bus and utility feeders, validate our designs.**

***Index Terms*—hosting capacity, voltage sensitivity, deep learning, data-driven method, long short-term memory (LSTM), spatial-temporal correlation, distributed energy resource**

I. Introduction

A rising number of renewable energy-based distributed energy sources (DERs) have been deployed in the distribution grid. The extensive use of DERs has many benefits, including regulating voltage profiles and lowering line loss. While this is happening, challenges also arise in the distribution grid due to the increasing penetration level. The widespread DERs, such as residential photovoltaics (PVs), inevitably affect the load shapes, voltages, fault current profiles, etc. Therefore, in solar-rich feeders, distribution system operators (DSOs) currently deal with overvoltage conditions instead of the low-voltage problems in the past. Solar energy frequently peaks at valley load for a feeder with many rooftop PVs. For example, the new PV deployment could cause a reverse current flow and impact the power supply of adjacent buses. Moreover, an overvoltage violation may occur at an upstream bus, even though the bus where power is injected does not see overvoltage. Managing solar planning and implementing appropriate control become emerging difficulties for grid planners and operators without a thorough understanding of the potential capacity to accept PVs. Therefore, hosting capacity analysis (HCA) is used to evaluate the distribution grid's capability for future sustainable operation.

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number DE-EE0008773.

Hosting capacity analysis determines the hosting capacity (HC) value which is defined as the maximum active power that DERs can safely inject into an existing distribution grid [1]. Traditional HCA methods typically conduct iterative power flow analysis or solve massive optimization problems to find the maximum point/value [2]. They target a snapshot of a feeder model with different load profiles, DER penetration, and other settings, thus quantifying the HC in this scenario. Moreover, some HCA methods focus on the change of HC over time. These methods provide insight into periodical HC patterns and increase the calculation accuracy for the time-varying feeder model [3]. However, the implementations of iterative optimizing functions are quite time-consuming for online/real-time analysis.

While model-based methods suffer from calculation burden, more historical data become easily acquired in distribution systems, which prompt data-driven analysis for system analysis and control [4]. To this end, we propose a learning-based problem formulation for HCA using historical measurements in distribution systems, shown in Fig. 1.

To accommodate time-series data and the underlying temporal correlation, we adopt the long short-term memory (LSTM) model [5]. The LSTM is an enhanced variant of the recurrent neural network (RNN) to capture the time-varying impacts on HCA. The RNN is widely used for time-series prediction but usually fails for a long sequence of data. The LSTM is modified from the RNN to capture the long-term memory [6]. However, the impact of spatial information cannot be embedded directly into the LSTM time-series prediction task. Specifically, the available spatial information includes locations of buses, current DERs, and potential DERs in the feeder. Since HC is often limited by violations of key criteria on adjacent or upstream buses, ignoring spatial correlation restricts the accuracy of HC determination.

Thus, to integrate the spatial and temporal correlations for HCA, we develop the basic LSTM to the Spatial-Temporal LSTM (ST-LSTM) model [7]. Specifically, our first contribution is the design of the dual forget gates in the structure of ST-LSTM. The gate function is the most crucial design in the

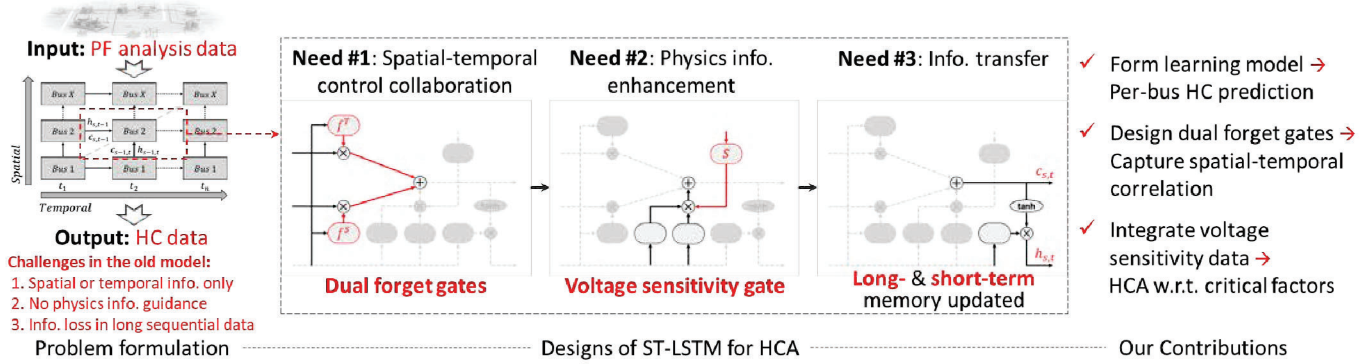


Fig. 1: Overview of the proposed ST-LSTM deep-learning model for hosting capacity analysis.

LSTM to capture temporal correlation, and the 'forget gate' is the most powerful gate among all the gates [8]. Therefore, we modified the forget gate to dual forget gates, memorizing temporal and spatial information in parallel.

Second, in the simulation, we found that voltage sensitivity data plays an important role in determining the HC because it is related to the voltage violations constraint, which is one of the most critical limits for deploying DERs. Specifically, [9] concludes that the voltage increase brought by the installed DERs is a severe problem for the grid. Therefore, we design the voltage sensitivity gate to improve the accuracy further. This gate can work with the input gate, another essential gate in the LSTM cell [10], to filter the inputs.

The remainder of this paper is organized as follows. Section II states the problem formulation. Section III demonstrates the spatial-temporal integration that converts the basic LSTM into the ST-LSTM as a physical embedding. Section IV shows the voltage sensitivity analysis that is highly correlated with HC and further leverage the information to enhance the ST-LSTM. Section V provides the numerical validation using different models. Conclusion and discussions are in Section VI.

II. Problem Formulation

A. Problem Statement

Unlike static snapshot hosting capacity, we focus on dynamic hosting capacity analysis (HCA), which helps the distribution system operators enhance hosting capacity determination, facilitate optimal control, and dispatch DERs in real time. We formulate HCA as a regression problem that learns a mapping from the power system data to the per-bus HC values. Given a scenario, learning-based HCA finds the maximum safe injection at one bus with fixing injections of all the other buses, which is the per-bus HC of this bus [11].

In the offline training, the input is the historical time-series data, including the voltage magnitude, voltage angle, load profiles, PV profiles, etc. Accordingly, the per-bus HC data is the desired output, also generated from the simulation. This approach allows our learning model to capture the HC-related information among detailed feeder settings. Subsequently, the learned model will give real-time predictions on the HC values with respect to the new inputs. Meanwhile, the model will be online updated based on the new inputs for better predictions.

B. Mathematical Definition for Learning-based HCA

With the historical data, we want to learn a regression model $f : X \rightarrow H$, where the input X is the power flow data with respect to time t. The learned model will predict the future hosting capacity value based on a new input. The deep learning-based HCA is defined as follows.

The offline training to learn $f : X \rightarrow H$:

- Input: time-series historical power flow data $X(t)$, where $x_{s,i}(t)$ is the value of the i-th power system feature component of bus s at time t.
- Output: time-series historical HC values $H(t)$, where $h_s(t)$ is the HC value of bus s at time t.

The forecasting based on $f : X \rightarrow H$ and online updating:

- Input: new time-series power flow data $X(\tau)$, where $x_{s,i}(\tau)$ is the value of the i-th power system feature component of bus s at time τ.
- Output: predicted time-series HC value $H(\tau)$, where $h_s(\tau)$ is the HC value of bus s at time τ.
- Online update: The new time-series power flow data $X(\tau)$ will update the model $f : X \rightarrow H$ and have a better prediction performance. If the spatial components, equipment settings, and availability of the system change in the feeder, the online update allows the model to learn the new pattern of the dynamic hosting capacity.

III. Spatial-Temporal Long Short-Term Memory

A. Recurrent Structure for Sequential Prediction

We use the Recurrent Neural Network (RNN) to consider the hosting capacity value changes over time in learning. A basic RNN is a feed-forward neural network. It is constructed by a series of identical units, known as 'RNN cells' [12]. The input of an RNN cell includes the output from the last cell, the hidden state h_{t-1}, and the input vector for this cell, which is denoted as x_t. Using these recurrent cells, an RNN can learn sequential information. Each cell has an output h_t as

$$h_t = \phi(W([h_{t-1}, x_t]) + b), \tag{1}$$

where $\phi(\cdot)$ is the nonlinear activation function.

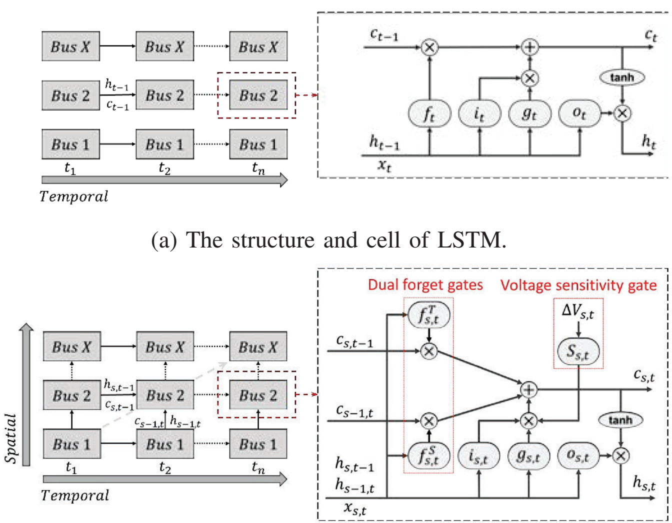
An RNN cell only has the hidden state h_t, which can capture short-term information. For the long sequence data, one of its variants, namely long short-term memory (LSTM), introduces the cell state c_t. The cell of LSTM uses three gates to update both h_t and c_t, memorizing the long-term information. Specifically, the forget gate, determining what information to forget, is defined as

$$f_t = \sigma(W_f([h_{t-1}, x_t]) + b_f), \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function used as the activation function. The sigmoid function non-linearly outputs between 0 and 1, where 1 means keeping the information, while 0 means abandoning the information. Similarly, this activation function is also used in other gates. The input gate i_t works with the candidate g_t to decide what information should be input. Besides, the output gate o_t controls the output of each cell. [8] The collaboration of all the gates will update the cell state c_t and the hidden state h_t as

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, \tag{3}$$
$$h_t = o_t \odot \tanh(c_t). \tag{4}$$

where c_{t-1} is the cell state of the last cell, the h_t is the hidden state forward-propagated to the next cell, and $\odot$ denotes the element-wise (Hadamard) product.



(a) The structure and cell of LSTM.

(b) The structure and cell of ST-LSTM.

Fig. 2: Structures of basic LSTM and proposed ST-LSTM. The dual forget gates and the voltage sensitivity gate in the ST-LSTM are highlighted in (b).

B. Temporal and Spatial Collaboration in ST-LSTM

A basic LSTM can powerfully process sequential temporal data, e.g., the historical time-series power flow data. Thus, this model can learn the HC changes over time under different time-series system conditions. Besides the temporal correlation, we also want to capture the spatial correlation in the hosting capacity analysis. Mutual effects among buses make the power grid extremely complicated. Therefore, we cannot independently consider one single bus and ignore the other buses in the analysis. However, an LSTM is a chain structure where one cell can only receive information from the last cell. Thus, the LSTM network cannot simultaneously process two sequences of data, i.e., temporal and spatial sequences. To overcome this limitation of the basic LSTM, we modify it to the ST-LSTM. The spatial sequences we used are the paths, and each path is from the feeder head to the lateral end. The reason for using paths is detailed in Section IV.

Specifically, we reconstructed the forget gate, the most critical gate in the LSTM, to dual forget gates, composed of the temporal and spatial forget gates. These two gates receive the cell states and hidden states from the last temporal cell and the last spatial cell, respectively, deciding what information should be memorized from two different dimensions. Mathematically, the new functional gates are

$$i_{s,t} = \sigma(W_i([h_{s,t-1}, h_{s-1,t}, x_t]) + b_i), \tag{5}$$
$$f^T_{s,t} = \sigma(W_{ft}([h_{s,t-1}, h_{s-1,t}, x_t]) + b_{ft}), \tag{6}$$
$$f^S_{s,t} = \sigma(W_{fs}([h_{s,t-1}, h_{s-1,t}, x_t]) + b_{fs}), \tag{7}$$
$$g_{s,t} = \tanh(W_c([h_{s,t-1}, h_{s-1,t}, x_t]) + b_c), \tag{8}$$
$$o_{s,t} = \sigma(W_o[h_{s,t-1}, h_{s-1,t}, x_t] + b_o), \tag{9}$$

where $h_{s,t-1}$ is the hidden state from the last time $t-1$, and $h_{s-1,t}$ is the hidden state from the last space $s-1$. $f^T_{s,t}$ represents the temporal forget gate, while $f^S_{s,t}$ is the spatial forget gate. The dual forget gates are highlighted in Fig. 2b.

Consequently, the new cell state, developed from (3), is

$$c_{s,t} = f^T_{s,t} \odot c_{s,t-1} + f^S_{s,t} \odot c_{s-1,t} + i_{s,t} \odot g_{s,t}. \tag{10}$$

where $c_{s,t-1}$ denotes the cell states from the last time while $c_{s-1,t}$ denotes the cell states from the last space. The multiplications of precious cell states and outputs from temporal and spatial forget gates determine which temporal and spatial information is needed, respectively. In this way, the ST-LSTM model captures in parallel the temporal and spatial correlations of power system data for learning-based HCA.

IV. Voltage Sensitivity Analysis

While the designed ST-LSTM model integrates inputs' temporal and spatial correlations, accurate learning needs specific guidance for HCA. In this section, we propose physics enhancement via voltage sensitivity analysis.

A. Voltage Sensitivity

When conducting the traditional hosting capacity analysis, we observe that the voltage violation constraint plays a crucial role, where a similar conclusion is found in [13]. High penetration of new PVs will bring overvoltage issues to the grid. Namely, voltage issues on sensitive buses will be a restraint on hosting capacity values. Thus, to analyze the impact of voltage violation limits on HC, we define the voltage sensitivity as the voltage changes of buses when deploying a fixed number of generations, e.g., $50kW$, at one bus. The voltage sensitivity data reveals specific physical correlations hidden behind the

978-1-6654-6441-3/23/$31.00 ©2023 IEEE

system topology for HC determination, but it remains difficult to embed this information into the learning process.

First, in the voltage sensitivity analysis, we found that different locations of the new PV have different impacts on the voltage changes of other nodes. Deploying the new PV at the lateral end will give and receive the most impact, experiencing and causing more significant shifts of bus voltage. Fig. 3 shows one example simulation in the utility feeder. This observation experimentally supports our design to use spatial information.

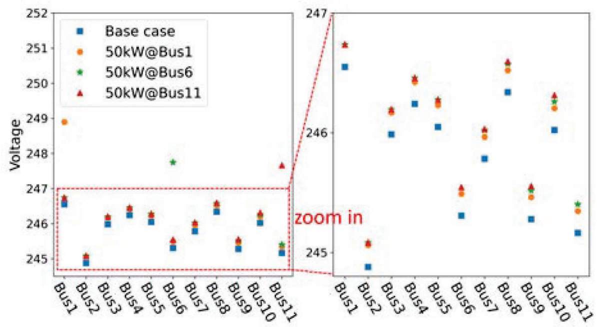


Fig. 3: Bus 1 is close to the three-phase main trunk, Bus 11 is the lateral end, and Bus 6 is in the path middle. The lateral end (Bus 11) brings the most significant voltage increases to other buses with the new $50kW$ PV deployment. Moreover, buses close to the lateral end significantly increase the voltage with the new PV deployment. Also, the data points of these buses are sparser, which indicates that these buses are more sensitive to the different locations of the new PV.

Second, we define the path as the feeder branch from the slack bus or the three-phase main trunk to each lateral end. We found that the new $50kW$ generation injections at one path will not bring significant voltage change to other paths compared to itself, as the example shown in Fig. 4. Therefore, supported by the independence property of the paths, it shows that the use of paths is meaningful.

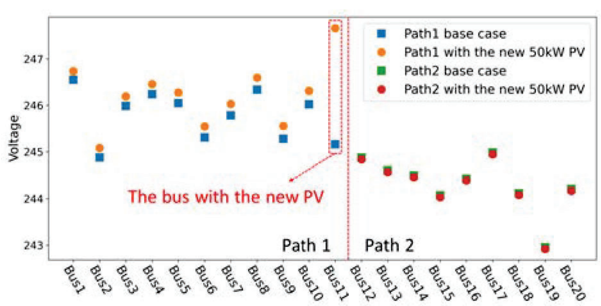


Fig. 4: Paths 1 and 2 are connected to the same three-phase transformer (different phases). With the new $50kW$ PV deployed at Path 1, bus voltage changes are larger here than at Path 2.

Third, we propose to guide the learning process with the voltage sensitivity data, which covers the hosting capacity-related information. Specifically, we use the average sensitivity data to incorporate the new injections' impact on the voltage changes of other buses based on locations. The number of average voltage changes for the other buses on the path to which bus s belongs is what we refer to as the average voltage sensitivity of bus s. Moreover, if we use voltage changes information considering only neighboring nodes, the resulting HC values have lower accuracy. Thus, we utilize the average bus voltage change of all bus nodes to represent each bus's influence on the overall path.

B. Sensitivity Information Enhancement in ST-LSTM

We design a voltage sensitivity gate, highlighted in Fig. 2b, to collaborate with the input gate in the proposed ST-LSTM. It uses voltage sensitivity information to control what new information can be added to the long-term memory c_t. Mathematically, the voltage sensitivity gate function is

$$S_{s,t} = \sigma(W_d \Delta V_{s,t} + b_d), \tag{11}$$

where $\Delta V_{s,t}$ is the voltage sensitivity data described in Sec. IV-A. Similarly, the sigmoid activation function $\sigma(\cdot)$ determines what information needs to be memorized. Consequently, the cell state function updated from (10) is

$$c_{s,t} = f^T_{s,t} \odot c_{s,t-1} + f^S_{s,t} \odot c_{s-1,t} + i_{s,t} \odot S_{s,t} \odot g_{s,t}. \tag{12}$$

Without modifying the output gate, the function of $h_{s,t}$ stays the same as (4), denoted as

$$h_{s,t} = o_{s,t} \odot tanh(c_{s,t}). \tag{13}$$

V. Numerical Results

A. Dataset Preparation in Distribution Grids

We use CYME to generate time-series data for validating our learning-based model. CYME is a power engineering simulation software by EATON that can be used to analyze power systems [14]. In CYME, the Load Flow module can provide power flow analysis, and the Integration Capacity Analysis module can calculate the HC. The Load Flow module generates voltage magnitude, voltage angles, load profiles, PV profiles, etc., which is the input of our model. The Integration Capacity Analysis module can set different constraints, including voltage violation constraints, thermal loading constraints, etc. Then, the software will progressively add PV and run power flow simulations until violations of one or more operation standards appear. Our learning-based model can naturally consider these constraints by using simulated HC data as training labels. We conduct the validations on the IEEE 123-bus feeder and an Arizona high penetration utility feeder. In the simulation, all deep learning models have the same hidden size of 300. The training and validation/testing datasets are generated from CYME with power system data as features and hosting capacity data as labels.

B. Evaluation Metric for Learning-based HCA

To measure the learning performance of the computed HC, we use percentage error as the evaluation metric. It gives an intuitive comparison between the simulated HC and the calculated HC. Specifically, simulated HC is the simulation result from CYME, and the calculated HC is the output of the deep learning models. The percentage error denotes

$$Percentage\ error = \frac{|calculated\ HC - simulated\ HC|}{simulated\ HC}$$

C. Improve Performance by Spatial-temporal Collaboration

This section validates the performance of the basic LSTM and the Spatial-Temporal LSTM model on the IEEE 123-bus feeder and the utility feeder. To show the performance of dual forget gates, we conduct time-series simulations to prepare data and implement the learning algorithms on the longest path for the IEEE 123-bus feeder and one zone in the utility feeder containing 12 paths. Table I shows that the ST-LSTM performs better on both feeders, which is due to integration of spatial and temporal correlations in power system data.

TABLE I: Percentage errors of learning models on feeders.

Test scenario	Temporal LSTM	Spatial LSTM	ST-LSTM
IEEE 123-bus	252%	22%	**5%**
Utility feeder	178%	12.2%	**7.9%**

D. Improve Performance by Voltage Sensitivity Information

In addition to dual forget gates for spatial-temporal correlations, the voltage sensitivity gate is designed to consider critical factors of HC determination. Therefore, this section validates the performance of the voltage sensitivity gate of ST-LSTM by showing the percentage error of the ST-LSTM with and without voltage in the sensitivity gate.

Fig. 5 shows the detailed experiment results on one example path in the utility feeder. We use three different models: the basic LSTM, the basic ST-LSTM, and the ST-LSTM with the voltage sensitivity gate. Since the sensitivity data enhances the consideration on other buses, the computation accuracy is not improved for every single bus, but the average accuracy increases on the entire path. Table I shows the percentage error of different models on the utility feeder. Notice that ST-LSTMs with the voltage sensitivity gate have lower percentage errors than the basic ST-LSTM because of the guidance of the voltage sensitivity data. We also use different approaches to embed the voltage sensitivity data. Approaches that consider overall voltage sensitivity, including the average and the standard deviation (STD) of voltage changes, perform better than those that only consider the voltage changes on adjacent buses.

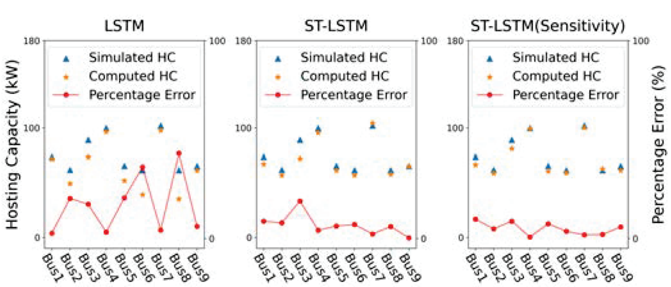


Fig. 5: The experiment results of three different models on one path in the utility feeder.

TABLE II: Percentage errors of ST-LSTM without and with the voltage sensitivity gate on the utility feeder.

Test scenario	Without sensitivity	Adjacent impact		Overall impact	
		Previous bus	Next bus	Average	STD
Utility feeder	7.9%	7.3%	7.1%	**5.9%**	6.2%

VI. Conclusion

This research develops a deep learning-based model called Spatial-Temporal LSTM to determine the dynamic hosting capacity in distribution grids. This paper makes three significant contributions. First, this research proposes a learning-based HC determination problem formulation. The LSTM is used to capture the HC changes over time. Second, this paper modifies the forget gate in the LSTM to dual forget gates to create a correlation between temporal and spatial sequences because of the forget gate's outstanding significance in the LSTM. Third, this paper improves the ST-LSTM using the voltage sensitivity gate. This gate implements the outcome of voltage sensitivity analysis. The proposed ST-LSTM significantly improves HC determination accuracy, as shown by comparing the basic LSTM and the ST-LSTM on IEEE example feeders and an Arizona high penetration utility feeder.

References

[1] J. Wu, J. Yuan, Y. Weng, and R. Ayyanar, "Spatial-temporal deep learning for hosting capacity analysis in distribution grids," *IEEE Transactions on Smart Grid*, vol. 14, no. 1, pp. 354–364, 2023.
[2] J. Yuan, Y. Weng, and C.-W. Tan, "Determining maximum hosting capacity for PV systems in distribution grids," *International Journal of Electrical Power and Energy Systems*, vol. 135, 2022.
[3] M. Rylander, J. Smith, and L. Rogers, "Impact factors, methods, and considerations for calculating and applying hosting capacity," *Electric Power Research Institute*, 2018.
[4] H. Li, Y. Weng, E. Farantatos, and M. Patel, "A hybrid machine learning framework for enhancing pmu-based event identification with limited labels," in *International Conference on Smart Grid Synchronized Measurements and Analytics*, 2019, pp. 1–8.
[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[6] Z. Zhang and M. Wu, "Locational marginal price forecasting using convolutional long-short term memory-based generative adversarial network," in *IEEE Power Energy Society General Meeting*, 2021, pp. 1–5.
[7] J. Liu, A. Shahroudy, D. Xu, A. C. Kot, and G. Wang, "Skeleton-based action recognition using spatio-temporal LSTM network with trust gates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 3007–3021, 2017.
[8] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2016.
[9] S. M. Ismael, S. H. A. Aleem, A. Y. Abdelaziz, and A. F. Zobaa, "State-of-the-art of hosting capacity in modern power systems with distributed generation," *Renewable Energy*, vol. 130, pp. 1002–1020, 2019.
[10] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *International Conference on Machine Learning*, 2015, pp. 2342–2350.
[11] J. Smith, M. Rylander, and L. Rogers, "Integration of hosting capacity analysis into distribution planning tools," *Electric Power Research Institute*, 2016.
[12] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
[13] M. Zain ul Abideen, O. Ellabban, and L. Al-Fagih, "A review of the tools and methods for distribution networks' hosting capacity calculation," *Energies*, vol. 13, no. 11, p. 2758, 2020.
[14] EATON, "CYME power engineering software." [Online]. Available: https://www.cyme.com/

978-1-6654-6441-3/23/$31.00 ©2023 IEEE