Deep Learning-Based Hosting Capacity Analysis in LV Distribution Grids with Spatial-Temporal LSTMs

by

Jiaqi Wu

A Thesis Presented in Partial Fulfillment
of the Requirement for the Degree
Master of Science

Approved April 2021 by the
Graduate Supervisory Committee:

Yang Weng, Chair
Raja Ayyanar
Elizabeth Cook

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

Nowadays, the widespread use of distributed generators (DGs) raises significant challenges for the design, planning, and operation of power systems. To avoid the harm caused by excessive DGs, evaluating the reliability and sustainability of the system with high penetration of DGs is essential. The concept of hosting capacity (HC) is used to achieve this purpose. It is to assess the capability of a distribution grid to accommodate DGs without causing damage or updating facilities. To obtain the HC value, traditional HC analysis methods face many problems, including the computational difficulties caused by the large-scale simulations and calculations, lacking the considering temporal correlation from data to data, and the inefficient on real-time analysis. This paper proposes a machine learning-based method, the Spatial-Temporal Long Short-Term Memory (ST-LSTM), to overcome these drawbacks using the traditional HC analysis method. This method will significantly reduce the requirement of calculations and simulations, and obtain HC results in real-time. Using the time-series load profiles and the longest path method, ST-LSTMs can capture the temporal information and spatial information respectively. Moreover, compared with the basic Long Short-Term Memory (LSTM) model, this modified model will improve the performance in the HC analysis by some specific designs, which are the sensitivity gate to consider voltage sensitivity information, the dual forget gates to build spatial and temporal correlation.

ACKNOWLEDGEMENTS

First, my deepest thank goes to Dr. Yang Weng. Under his guidance and supervision, problems and confusion in the thesis were solved one by one, and the thesis was finally completed successfully. His profound knowledge and rigorous style of study have benefited me greatly and have contributed to the successful completion of this thesis. I would like to express my heartfelt thanks to him!

Second, I want to express the depth of my gratitude to my committee members, Dr. Raja Ayyanar and Dr. Elizabeth Cook for giving me instructive advice and their direct and indirect help.

Then, I would like to thank my group member Jingyi Yuan, who gave me lots of help and instruction on my thesis.

Finally, I would like to express my sincere thanks to my family and my friends, especially my mother and my friend Jiaxin Zheng. Without the support from them, I would not have made it so far.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Nowadays, a growing number of renewable energy-based distributed generators (DGs), especially photovoltaics (PVs), have been widely used in the low-voltage (LV) distribution grid. While the widespread use of DGs has many advantages, including voltage profile control, line loss reduction, and cost decrease, it also brings challenges to planning and optimizing the distribution grids. For instance, DGs may integrate into the distribution grid on feeder level or behind-the-meter. They are highly distributed, variable, and often beyond utility monitoring. It is currently difficult to receive timely status updates due to the lack of infrastructure and proper analytics. When penetration level is substantial, DGs inevitably change load shapes, voltage, and fault current profiles, and other operating conditions. For example, instead of low voltage issues in traditional radial feeders, now grid operators face overvoltages at solar-rich feeders. For communities with lots of roof-top PVs, solar energy often peaks at valley load. The net load then turns negative and introduces reverse power flow and overvoltages at the feeder end, causing malfunction of voltage regulators and protection coordination. Without proper situational awareness, grid planners and operators find it challenging to handle capacity planning and take appropriate control actions.

To assess the power network, the concept of hosting capacity (HC) is proposed. HC is denoted as the maximum amount of DGs that an existing distribution grid can accommodate without causing technical problems or requiring changes to power system facilities [1]. [2] provides a systematic and comprehensive introduction to the research, development, evaluation, and enhancement of HC. [3] offers a classification of methods for quantifying HC and accomplishes a detailed review of three different methods for calculating HC. [1] not only does a similar review work but also summarized current tools using in HC analysis. All the works about HC illustrate that this concept has been accepted by many researchers and engineers to evaluate the distributed power system. To obtain the HC value, some different methods are applied.

In [3], current HC analysis methods are classified into four categories. The deterministic method [4, 5, 6] is to calculate the HC at the worst-case scenario. Without considering the complex scenario of a real power system, the obtain HC value is underestimated and lacks certainty. The stochastic method [7, 8, 9] is to use the Monte-Carlo-based technique for uncertainties in power system, including random placement and sizing of PV units and loads. This popular method also has some drawbacks, e.g., the computational time, and will increase the complexity of the power system and uncertainty parameters. And this method can not assess the time-related factors that occurred in the real power system. The time series method [10, 11] is to apply generated time-series data in simulation and also can be treated

as a deterministic method. It is a very time-consuming method that requires a lot of simulation. Although this method can consider the time-series factor, in some scenarios, it is unable to consider the model changes between two very long time slots. The optimization-based method targets maximizing the active power injection of DGs into the distribution network while making sure that the distribution network operational limits are satisfied. This method needs to do many iterations to obtain an optimal solution.

Obviously, all of the current methods are based on a large-scale simulation and calculation to improve the accuracy, and the complexity will significantly increase accompanied by the uncertainty in the system. Additionally, most of the methods ignored the time factors in HC analysis. Moreover, all of the above methods are not designed for real-time analysis, instead, they are focusing on the calculation of the HC of the system generally, rather than the HC value based on different load profiles.

Based on the consideration of solving these problems, we choose to use a machine learning-based method for HC analysis. The input of our model is the historical time-series power system data. Then we can obtain the HC as the corresponding output in real-time, using the trained model. To guarantee the accuracy of the calculation, choosing a proper machine learning model is very important.

Nowadays, variant machine learning methods have been already widely used in different power system analyses to solve specific problems. [12] developed a

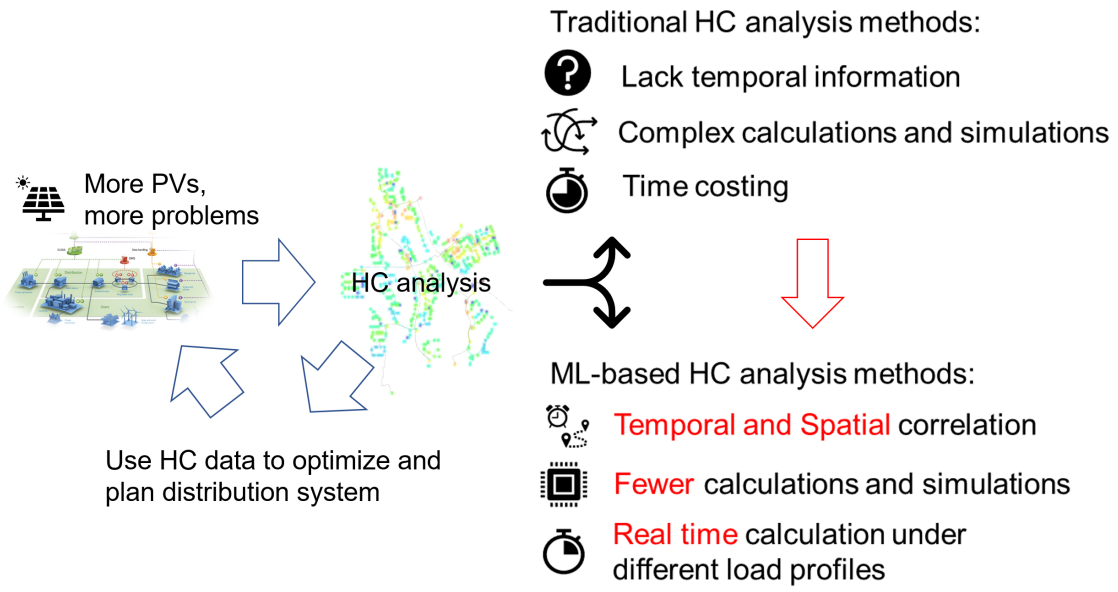


Figure 1.1: Overview

new scheme to mitigate the battery energy storage systems (BESS) capacity problem through Monte-Carlo tree search-based reinforcement learning (MCTS-RL). [13] used a data-driven regression approach to train a linearized power flow model. In the recent HC analysis study, some machine learning methods have been applied as well. [14] proposed a machine learning-based method, in particular support vector machines (SVM), to classify the grids with a high penetration of DGs, using sample data. This method can significantly better reflect domain expert assessments. There are also some machine learning-based methods used in HC analysis. In [15], the author introduced a machine learning-based method, the Bayesian Optimization, to do HC analysis. It can prove that data-driven learning-based methods for HC analysis in the power system are feasible. However, this model is

not good at considering the temporal correlation in HC analysis.

To choose a proper model for machine learning-based HC analysis, we can focus on the features of power systems. The power system is a well-formed sequential model which contains temporal and spatial sequences. The Recurrent neural network (RNN) is proposed to solve the sequential problems, and one of its variants, the LSTM, has a better performance in practice compared with basic RNNs. Traditional LSTMs have some drawbacks. The most important one is that it can only consider one type of sequential information at the same time. However, in the HC analysis, spatial sequence, including electrically and geographically relative locations, and temporal sequence, containing time-series load profiles, both play important roles in HC analysis. The motivation to consider both spatial and temporal correlation existed in the special power system requires us to develop the traditional LSTM to the ST-LSTM. ST-LSTM models have been used in many situations. [16, 17, 18] used this method to do prediction on people's next destination. [19] used the LSTM model to mimic the physical action of a person. [20] introduced a tree-structure-based LSTM model, which inspired us to apply a similar method in our ST-LSTM model.

For my new HC analysis method, first, given historical time-series power system data, this deep learning-based approach can compute the HC in real-time. Second, the longest path method will build a relatively simple spatial sequence for the complex power system. Third, the new ST-LSTM algorithm builds a tempo-

ral and spatial correlation, the designed dual forget gates decide what information needs to be memorized, and the sensitivity gate can filter the input information based on the voltage sensitivity analysis.

Chapter 2

HC ANALYSIS METHOD

2.1 Comparison Between Traditional and Deep Learning-based HC Analysis

Traditional HC analysis methods are based on either extremely large datasets for instants the time-series method, or repeat simulation, for instants the optimization-based method. Nevertheless, they are not used to calculate HC in real-time. Our target is to calculate the HC in real-time with high accuracy.

HC data is very sensitive to temporal and spatial information, both of which can be treated as sequential data. In the HC analysis deep learning framework, we use these two kinds of sequential labels, temporal and spatial. The temporal label means the time interval between two data, and the spatial label means the location of PV generators. To be more specific, temporal information means the time series data, and spatial information means the all the buses in the tree-structured power network can follow a well-setting order, which will be talked about in detail in the following subsection. HC number is relaid on spatial and temporal information, both of which are following a well-formed order. Specializing in capturing both long-term and short-term sequential information, LSTM is chosen to be our deep learning model.

2.2 Voltage Sensitivity Analysis

Moreover, in our analysis, we noticed that when adding PV generators at one selected bus, the limit of voltage violation for other buses plays a crucial role in HC analysis. Based on this, we use this voltage sensitivity data as a part of the input of the sensitivity gate.

The voltage sensitivity of a local bus i is defined as how much impact it has on the voltage of the other buses when adding a generator with unit generation. [21] emphasizes that according to the literature, voltage rises at load bus bars are a serious limiting factor when installing DGs. In [1], the author emphasizes that over-voltages and under-voltages are the main limiting factor in determining HC.

Chapter 3

BASIC LSTM

Targeting on capturing temporal correlation among data, a sequence-sensitive machine learning method is used in our work. To achieve such learning, the mapping from historical data on different input features to the hosting capacity is observed to be highly nonlinear, and deep learning is a promising method to deal with such non-linearity. However, to describe the time-series in power systems, it is crucial for memory to exist between neighbor neurons to capture the time dependence of the physical process. The RNN is designed for this purpose. In such a model, RNNs need some past context to predict the present output, but the issue is how much. It seems that sometimes, less context is sufficient, but it is entirely possible that the gap between the relevant information and the point where it is needed becomes very large. This is the issue of long-term dependencies, which is almost impossible to capture using classical linear or Markov-based systems[22]. In theory, RNNs should be capable of learning memories of arbitrary length, but in practice, they don't seem to learn them.

In the mid-90s, a variation of the recurrent net with so-called Long Short-Term Memory units, or LSTMs [23], was proposed as a solution to the vanishing gradient problem. The main idea is that long memories were observed to be missing in the

application results from RNN. However, a power system usually has long-term patterns, ranging from daily, weekly, and monthly, to seasonal periodicity. For this purpose, LSTM aims at remembering these long-term relationships, favorable for power grids. Therefore, using LSTM for calculating hosting capacity is vital for achieving high accuracy when compared to other learning methods.

3.1 Illustration of the Basic LSTM

An RNN is composed of a series of units which are named cells, and each cell is constructed as a hidden single-layer neural network in the most basic RNN. Each RNN cell has two input parts, one is the output from the last one and the other is a piece of data from the dataset. In this way, these repeating cells can transfer the information and the network will be trained recurrently. RNNs can be algebraically described as

$$h_t = \phi(W[h_{t-1}, x_t] + b), \tag{3.1}$$

where ϕ is the activation function.

Though RNNs are designed to process the sequential data, it has drawbacks, one of which is that RNNs are unable to transfer information correctly between two cells with relatively more intervals gap. To solve this problem, LSTM [23] is developed from traditional RNN and capable of learning short-term and long-term dependencies. This gradient-based algorithm can also avoid gradient exploding and vanishing which are also significant drawbacks of RNNs.

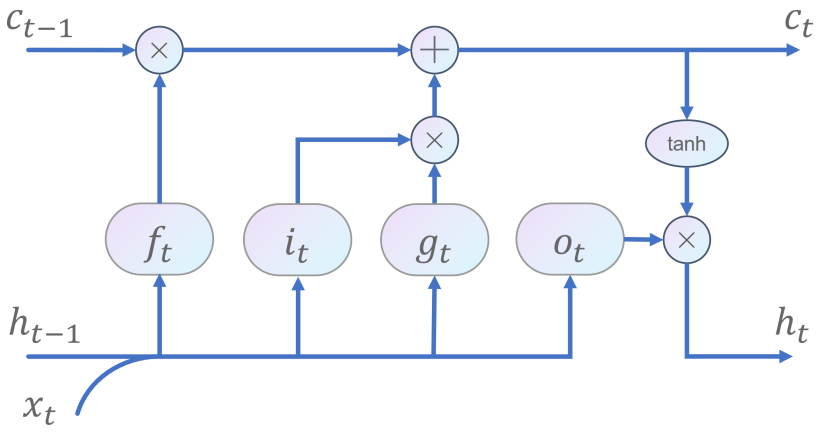


Figure 3.1: The Basic LSTM Cell

Aiming on addressing these problems, LSTM has become an effective model for sequential prediction problems. Similar to the RNNs, an LSTM network also has a series of connected cells, shown in Fig 3.1. Differently, the highlight of this state-of-the-art algorithm is the design of the three gates to control the information stream. The basic gate functions to update traditional RNNs are

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{3.2}$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \tag{3.3}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \tag{3.4}$$

where i_t, f_t and o_t represent the input, forget and output gates of the t-th cell, respectively. The input gate decides what information should be input. The forget gate decides what information should be forgotten. The output gate decides what

information needs to be output. These three gates use the activation function sigmoid function, which is represented as $\sigma(\cdot)$. The sigmoid function, which is shown in the Fig 3.2, is

$$Sigmoid(x) = \frac{1}{1 + e^{-x}}. \tag{3.5}$$

This nonlinear function, $\sigma(\cdot)$, represents a sigmoid layer that can map the values between 0 to 1, where 1 means to keep the information while 0 means to forget the information.

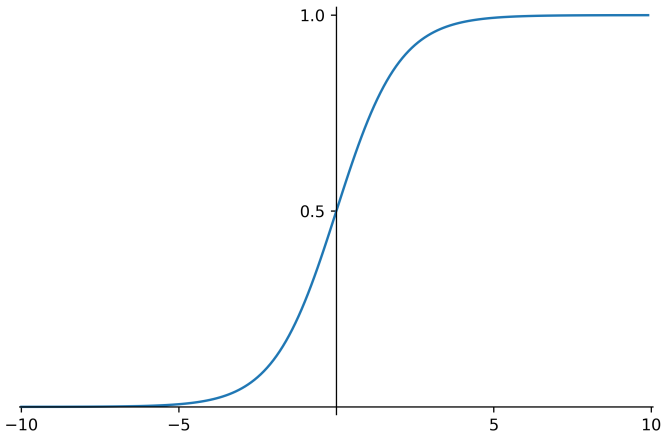


Figure 3.2: Sigmoid Function

Similar to the basic RNN function, one of the component of the new cell state, the candidate g_t, is

$$g_t = \tanh(W_c[h_{t-1}, x_t] + b_c), \tag{3.6}$$

where the activation function $\tanh(\cdot)$, which is shown in Fig 3.3, is

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{3.7}$$

This activation function will increase the linearity of the model.

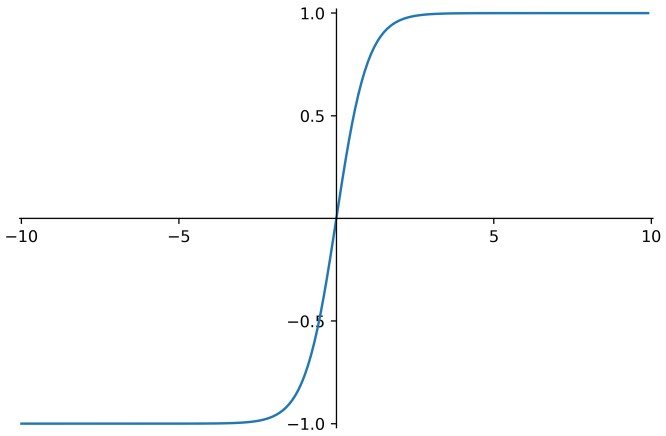


Figure 3.3: Hyperbolic Tangent Function

Each of the gates and the candidate is like a small neural network, in which x_t and h_t represent the input feature vector and the hidden output vector. W_i, W_f , W_o and W_c are the weights of each gate and b_i, b_f , b_o and b_c are corresponding biases. Fig 3.4 demonstrates the gate structure, using the input gate as the example.

The cell state updating function is

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, \tag{3.8}$$

where the former part is the previous cell state c_{t-1} that is controlled by the forget gate f_t, and the latter part is the input vector scaled by new candidate value

$$i_t = \sigma(W_i[h_{t-1},, x_t]) + b_i)$$

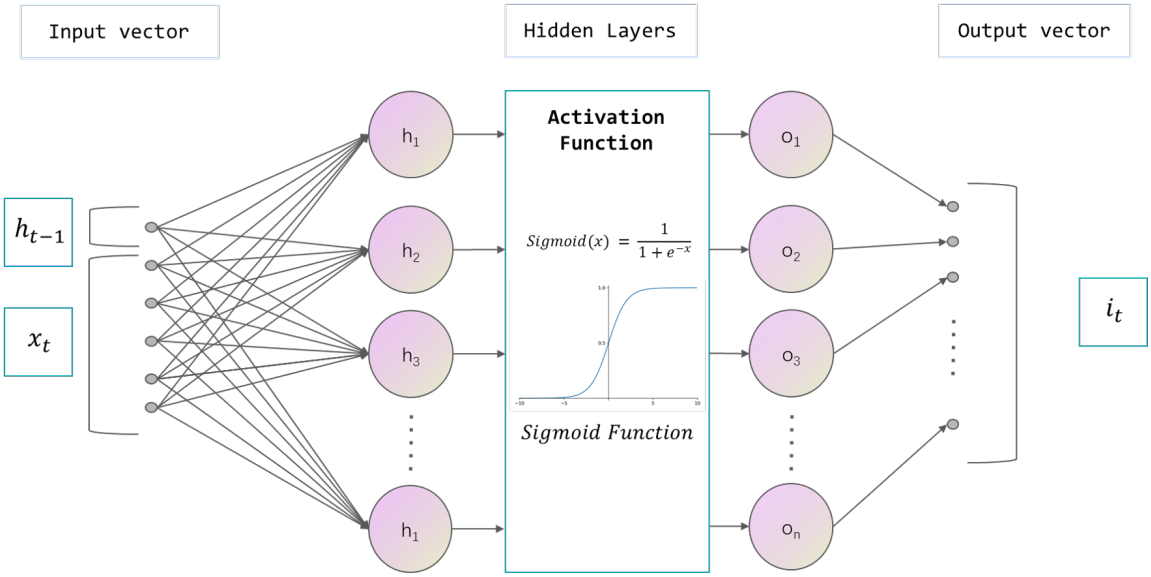


Figure 3.4: Gates in the Basic LSTM

which to decided how much of input to add state value. The two parts of this function will work together to update the cell state c_t .And $\odot$ represents the element-wise (Hadamard) product.

The output function is denoted as

$$h_t = o_t \odot \tanh(c_t), \tag{3.9}$$

where the h_t is the hidden state that needs to be transferred to the next cell. What is worthy of notice is the output h_t of the last cell is the final result we want to obtain, which is HC value in our problem.

Chapter 4

SPATIAL-TEMPORAL LSTM

Traditional LSTM models use cell state c_t to store long-term sequential information, streaming through the whole LSTM network, while hidden layer h_t process relatively short-term information. However, both c_t and h_t come from the same series of sequential data, which means a cell can only receive information from the last cell, and this repeating network is trained as a single chain. Thus, the LSTM network is unable to treat these two different types of sequential data at the same time. Our target is to calculate the HC in real-time using time series data. Furthermore, a power system is a complex network, which means geographical information plays an important role in HC analysis. The basic LSTM model is unable to build the spatial correlation among buses and temporal correlation among time-series datasets in the same model. To solve this problem and take spatial information into consideration, we need to update the basic LSTM to the ST-LSTM.

The structure of the basic LSTM and the ST-LSTM are shown in Fig 4.1 and Fig 4.2 respectively.

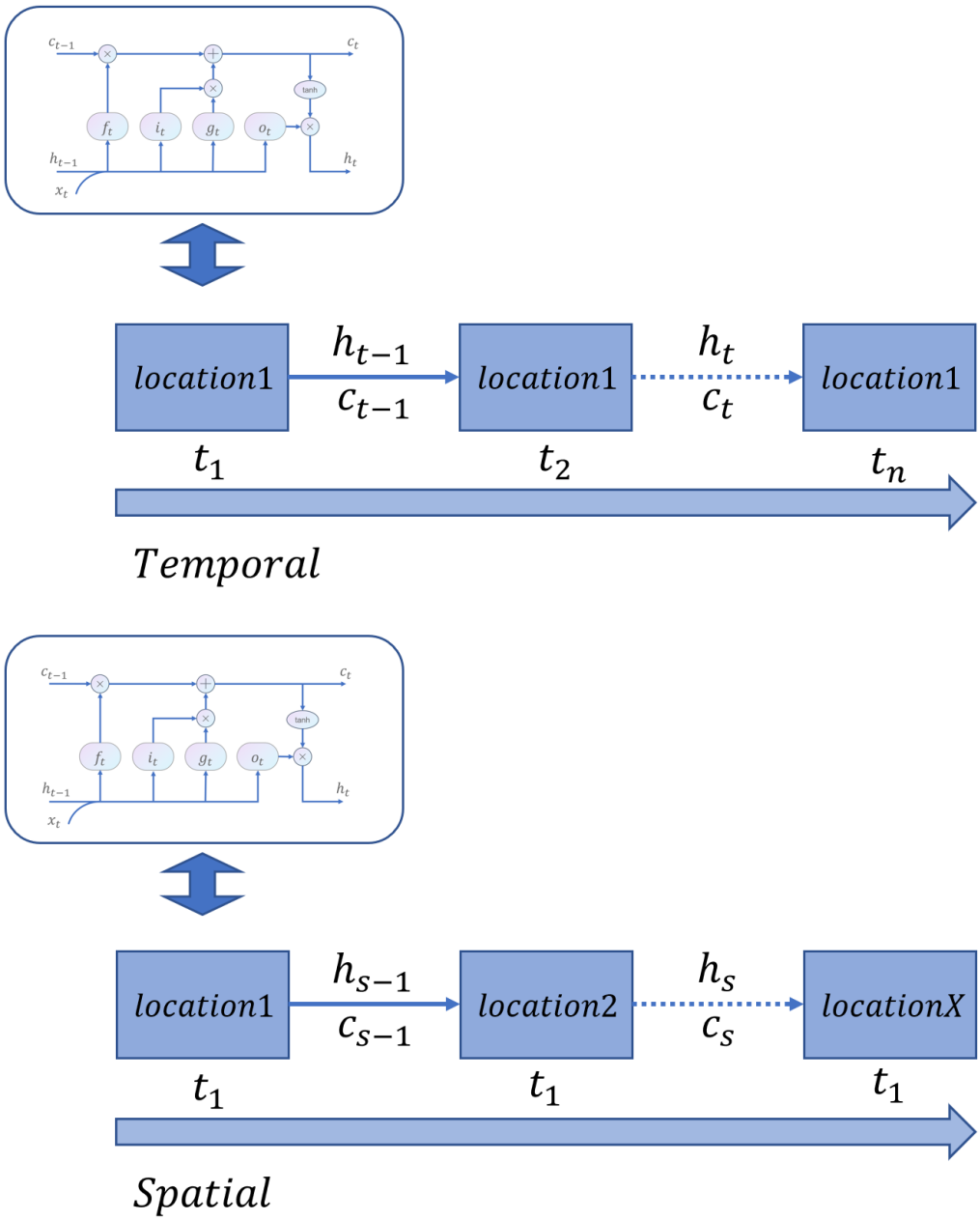


Figure 4.1: The Structure of the Basic LSTM

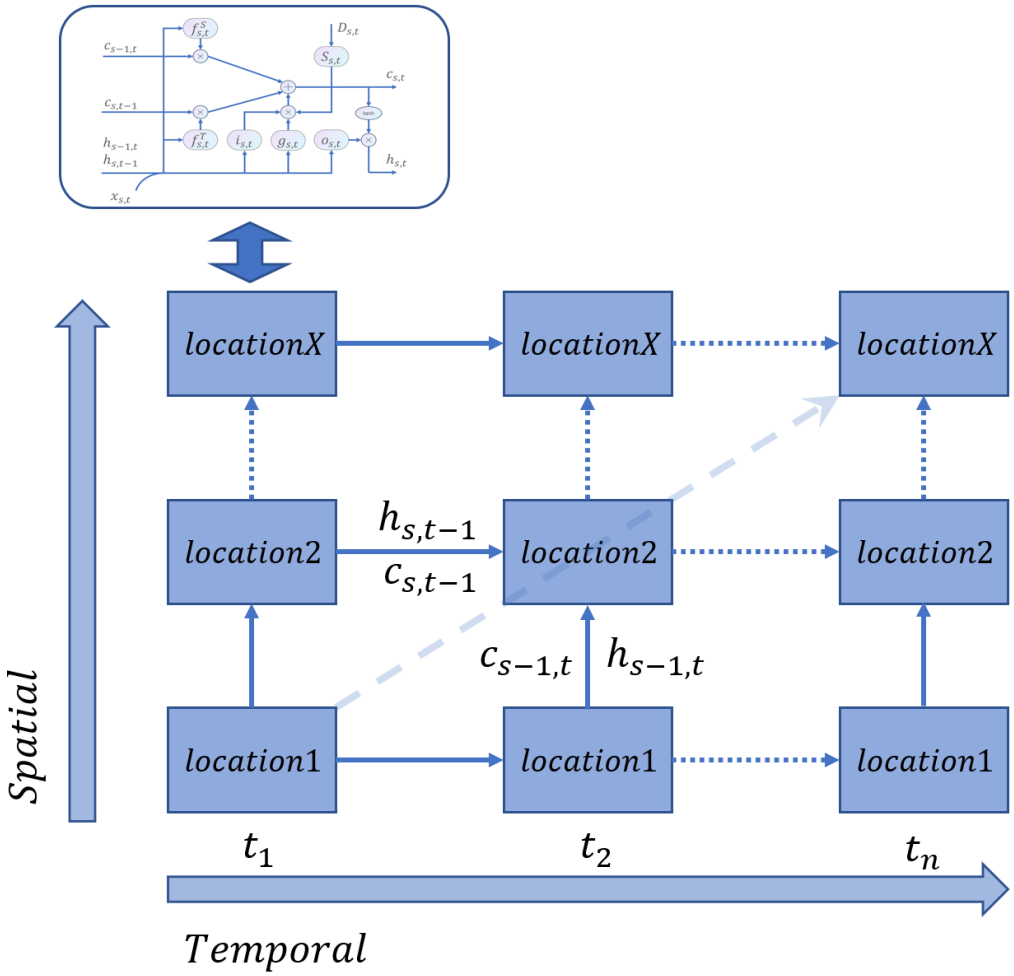


Figure 4.2: The Structure of the ST-LSTM

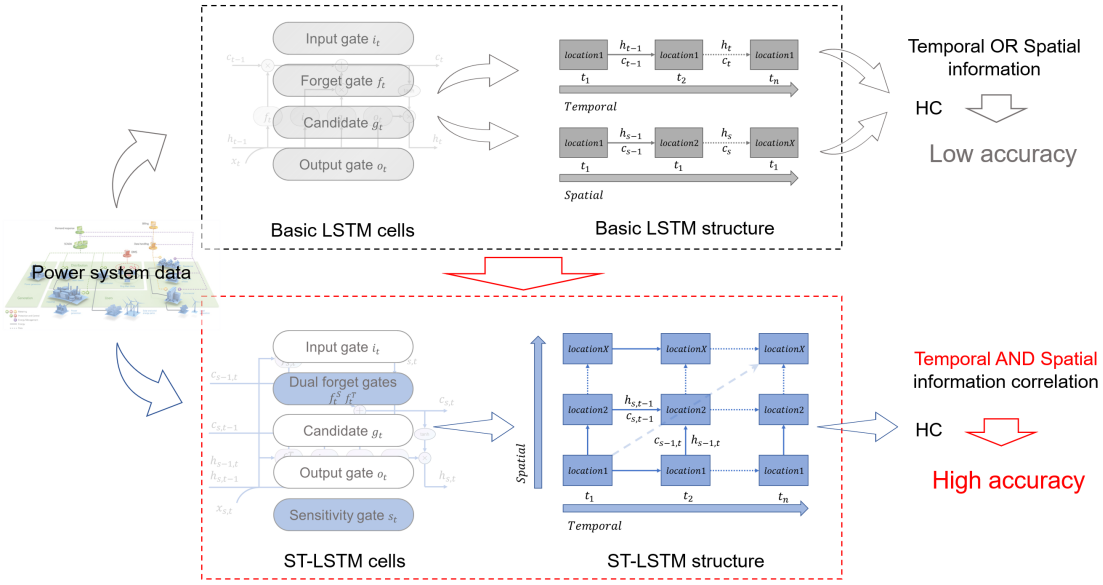


Figure 4.3: The Comparison Between the Basic LSTM and the ST-LSTM

The comparison between this two models is shown in Fig 4.3.

4.1 Design of the Dual Forget Gates

In order to process two sequences, specifically spatial and temporal sequence, in parallel in our deep learning neural network, we update the traditional LSTM algorithm to ST-LSTM. Fig 4.4 illustrates the updated ST-LSTM cell.

The new design of gates in the ST-LSTM are

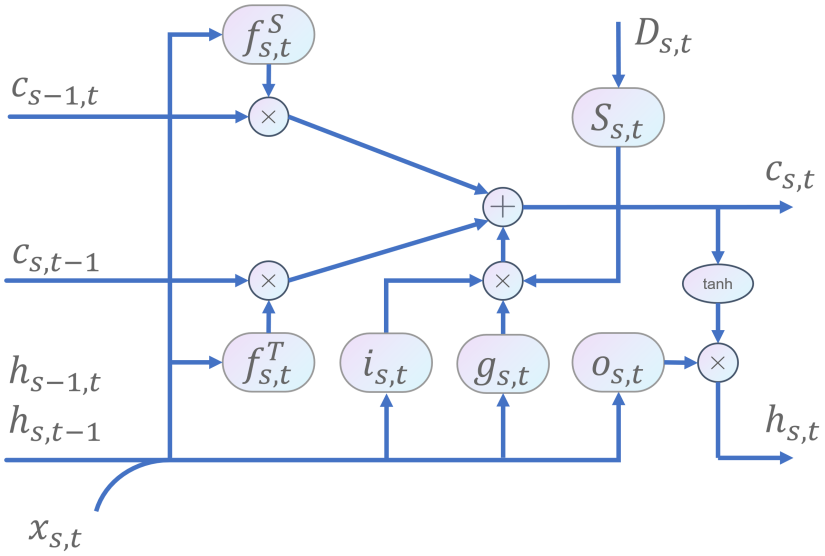


Figure 4.4: The ST-LSTM Cell.

$$i_{s,t} = \sigma(W_i([h_{s,t-1}, h_{s-1,t}, x_{s,t}]) + b_i), \tag{4.1}$$

$$f^T_{s,t} = \sigma(W_{ft}([h_{s,t-1}, h_{s-1,t}, x_{s,t}]) + b_{ft}), \tag{4.2}$$

$$f^S_{s,t} = \sigma(W_{fs}([h_{s,t-1}, h_{s-1,t}, x_{s,t}]) + b_{fs}), \tag{4.3}$$

$$g_{s,t} = \tanh(W_c([h_{s,t-1}, h_{s-1,t}, x_{s,t}]) + b_c), \tag{4.4}$$

$$o_{s,t} = \sigma(W_o[h_{s,t-1}, h_{s-1,t}, x_{s,t}] + b_o), \tag{4.5}$$

where $h_{s,t-1}$ is the hidden state from the last time and the $h_{s-1,t}$ is the hidden state from the last space. Fig. 4.5 demonstrates the gate structure, using the input gate as the example.

And more importantly, $f^T_{s,t}$ represents the forget gate based on temporal information, while $f^S_{s,t}$ represents the forget gate based on spatial information. These

$$i_{s,t} = \sigma(W_i[h_{s,t-1}, h_{s-1,t}, x_{s,t}]) + b_i)$$

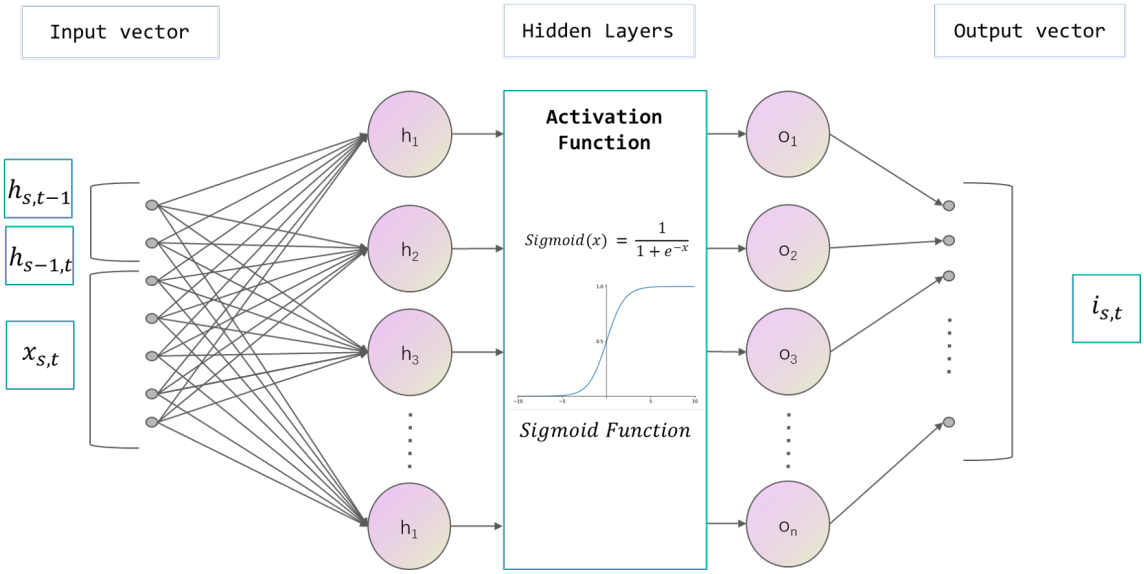


Figure 4.5: Gates in the Basic ST-LSTM

two gates, highlighted in Fig. 4.6, are the designed dual forget gates.

Because of the design of the dual forget gates, the cell state of this cell will be impacted by both temporal forget gate and spatial forget gate, and the new cell state function is

$$c_{s,t} = f^T_{s,t} \odot c_{s,t-1} + f^S_{s,t} \odot c_{s-1,t} + i_{s,t} \odot g_{s,t}. \tag{4.6}$$

We multiply the cell state $c_{s,t-1}$ in an element-wise way by our temporal-based forget gate $f^T_{s,t}$ to decide what temporal information needs to be memorized. Similarly, the element-wise multiplication of the cell state $c_{s-1,t}$ and spatial-based forget gate $f^S_{s,t}$ is to determine the keeping and forgetting of the spatial information.

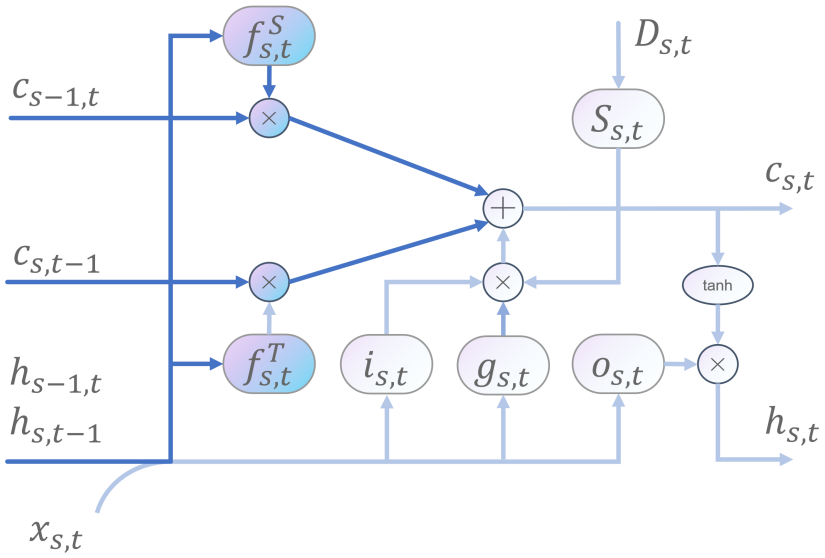


Figure 4.6: The ST-LSTM Cell with the Dual Forget Gates

The function of $h_{s,t}$ will be remained as

$$h_{s,t} = o_{s,t} \odot tanh(c_{s,t}). \tag{4.7}$$

4.2 Design of the Sensitivity Gate

We designed dual forget gates to decide which information should be memorized. To decide which information should be input, besides the current input gate, we designed a new gate, namely the sensitivity gate, which includes the distance information. The sensitivity vector is the input of the sensitivity gate, and it contains some different types of distance information. The most obvious information is the electrical distance, calculated based on the length of the transmission line. In this way, the model will reinforce the spatial correlation. The highlighted part in the Fig 4.7 is the designed sensitivity gate.

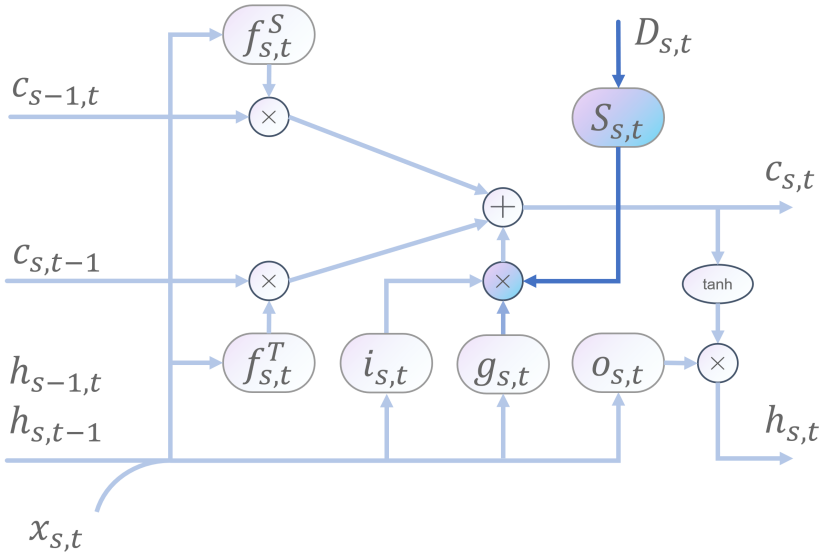


Figure 4.7: The ST-LSTM Cell with the Sensitivity Gate

The sensitivity gate is denoted as

$$D_{s,t} = \sigma(W_d \Delta d_{s,t} + b_d), \tag{4.8}$$

where Δd_t is the sensitivity vector.

In our analysis process, we found that HC is mostly limited by the voltage violation constrain. Thus, we use the result of voltage sensitivity analysis in our sensitivity gate. This voltage sensitivity is also based on the network structure and can be regarded as a kind of distance in a different sense.

The sensitivity gate, collaborating with the input gate, uses the voltage sensitivity information to control the input when adding new information into the cell state. It is to decide what new information should be added to update the cell state.

The updated cell state calculating function is

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

4.3 Information Flow

The information flow of the ST-LSTM can be illustrated in Fig 4.8.

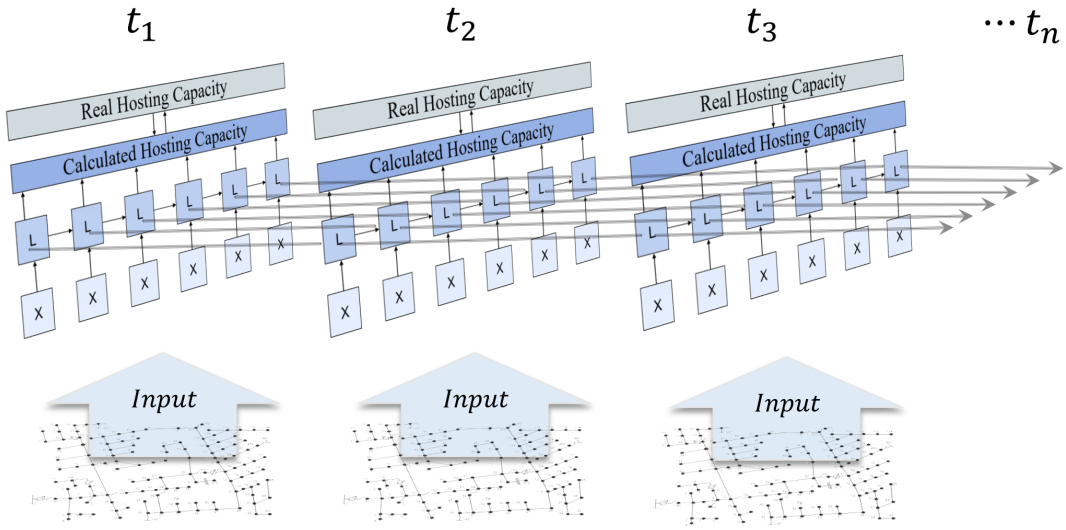


Figure 4.8: The Information Flow in the ST-LSTM

The dimension of each data is

$$x_{s,t} \in \mathbb{R}^n, \tag{4.9}$$

$$h_{s,t} \in \mathbb{R}^m, \tag{4.10}$$

$$c_{s,t} \in \mathbb{R}^m, \tag{4.11}$$

$$f^T_{s,t} \in \mathbb{R}^m, W_{ft} \in \mathbb{R}^{m\times(2m+n)}, \tag{4.12}$$

$$f^S_{s,t} \in \mathbb{R}^m, W_{fs} \in \mathbb{R}^{m\times(2m+n)}, \tag{4.13}$$

$$i_{s,t} \in \mathbb{R}^m, W_i \in \mathbb{R}^{m\times(2m+n)}, \tag{4.14}$$

$$g_{s,t} \in \mathbb{R}^m, W_c \in \mathbb{R}^{m\times(2m+n)}, \tag{4.15}$$

$$o_{s,t} \in \mathbb{R}^m, W_o \in \mathbb{R}^{m\times(2m+n)}, \tag{4.16}$$

$$\Delta d \in \mathbb{R}^k, \tag{4.17}$$

$$D_{s,t} \in \mathbb{R}^m, W_d \in \mathbb{R}^{m\times k}, \tag{4.18}$$

where m is the hidden layer size and n is the input vector size. $m = 1$ in our problem, because our required output is an HC value.

Chapter 5

SPATIAL SEQUENCE

5.1 Longest Path Method

Because of the complexity of the power system, it is difficult to put all the buses into a sequence. Thus, to convert the network into a spatial sequence, I choose paths in the network. For each time series, the ST-LSTM model will output the HC of all the buses in this path as a vector.

The longest paths are shown in Fig 5.1. This path contains two important regulators and node 67 which is one of the most critical nodes in HC analysis.

Besides the longest path, different paths will compose the whole network. Thus, we can extend the definition of the longest path to a path which is from the feeder head to the feeder end. Since we will get the HC of all the buses in the path when training it, theoretically, we can find a finite number of paths to cover the whole network. In this way, we are able to calculate the HC of all buses in the network.

5.2 Tree Structure-based Traversal

If we use all the buses at the same time, we need to decide the traversal order. Otherwise, the improper, even random, order will weaken the spatial connection between nodes, because we need to use the context information, which is $c_{s-1,t}$ and

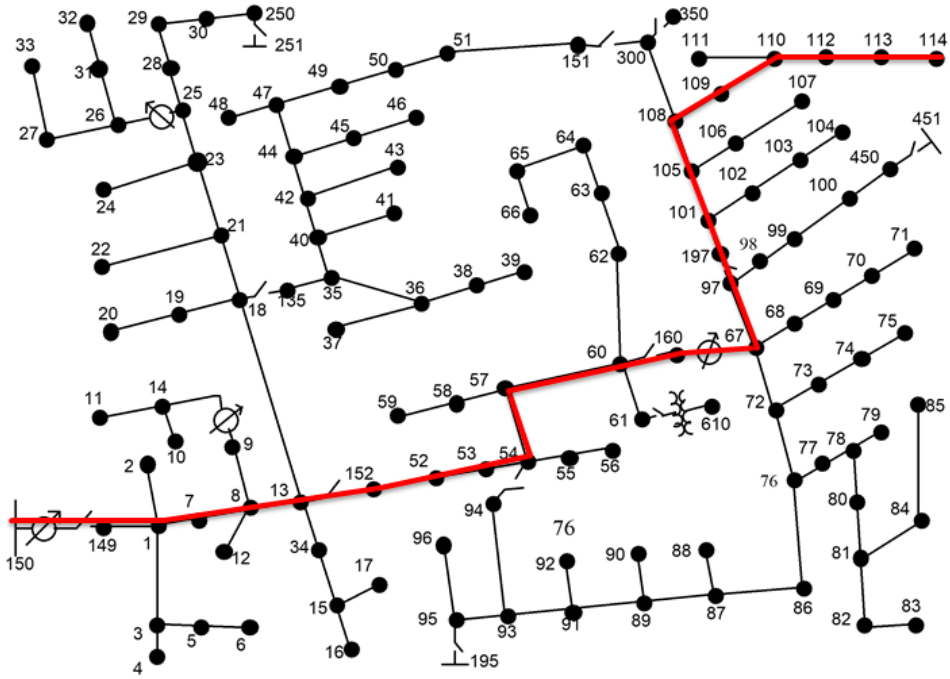


Figure 5.1: Path 1 in the IEEE 123-bus System

$h_{s-1,t}$.

The electric system can be treated as a tree structure network. Even though some power network names their buses following numerical orders, like IEEE 123 bus network, this kind of order can not represent an exactly correct relatively geographical relationship information among all the buses. For instance, in IEEE 123 bus network, node 33 and node 34 have an adjacent numerical order, but node 33 is at the feeder end, while node 34 is at the feeder middle, away from node 34 geographically. Additionally, in real power system networks, they are more likely to name their buses using character strings for classification purposes. Both of the two problems will weaken the spatial relationship if we import the dataset in

simple chain order.

To solve this problem, we plan to use a bi-direction tree traversal scheme. In this scheme, the traversal begins from the feeder head, which is treated as a root node, and visits all the buses in sequence. Once reaching feeder ends, which are treated as leaf nodes, the traversal will turn backward in the tree. Finally, the whole traversal will end at the feeder head. In this scheme, each node will be visited at least twice from two different directions. Thus, in our ST-LSTM model, one cell will receive $c_{s-1,t}$ and $h_{s-1,t}$ information from both of its neighbour, considering the adjacent information from both sides. Also, since this traversal will meet each node twice from a different direction, it can simulate the reverse power flow caused by the high penetration of PV generators.

Chapter 6

NUMERICAL RESULTS

Based on the comparison of basic LSTM and our ST-LSTM, we use the simulated data to support our design. To test the performance of the ST-LSTM model, we use both temporal sequence LSTM and spatial sequence LSTM as our baseline model.

6.1 Data Preparation

CYME is a power engineering simulation software by EATON which can be used to analyze power systems. In CYME, the Load Flow module can do power flow analysis and the Integration Capacity Analysis module in CYME can calculate the HC. We have 12 time-series models of the 123-bus network in CYME.

We use the result of Load Flow analysis as our input. Specifically, we use the bus voltage (p.u.), angles, and the thru power of each phase. For the input data, we use the Min-max normalization method, which means for each type of feature, the minimum value of that feature is converted to 0, the maximum value is converted to 1, and all other values are converted to a decimal number between 0 and 1. For the HC data, we directly use a selected constant to scale the value, which can allow us to retrieve the output to normal scale easily.

Specifically, simulated HC is the simulation result from CYME and the calcu-

lated HC is the output of the deep learning models.

6.2 Experiment Setup

As shown in Figure 4.3, the basic LSTM model can only flow one type of sequential information, either temporal sequence or spatial sequence. Thus, we used basic LSTM to test the performance using both of them.

For temporal sequence, we have IEEE 123-bus time series models. In each training step, we input the dataset from power flow analysis of all the time at one bus and are able to have the output as the HC data vector for all the time slots at this bus. The training sets are from 20 buses and the testing sets are from 4 buses.

For spatial sequence, we chose the longest path in the IEEE 123-bus network which is from the feeder head to the feeder end. This path contains 24 buses and we picked up their data from all the buses' data. Then, in each step, we put them in the network and obtain the HC data vector of all the buses at this time. The training sets are from 9 timeslots and the testing sets are from 3 timeslots.

For our ST-LSTM model, we did a similar thing as the spatial sequence LSTM. The difference is that, in each step, we use the outputs from the last time as part of the input. The training sets are from 9 timeslots and the testing sets are from 3 timeslots.

We use two criteria to evaluate the performance of the models. The first one is the Mean square error (MSE) criteria and the other one is the percentage error. The MSE is used by the model to do back-propagation and the percentage error can

give a direct comparison between simulated HC and the calculated HC.

6.3 Results and Comparison of Different Models

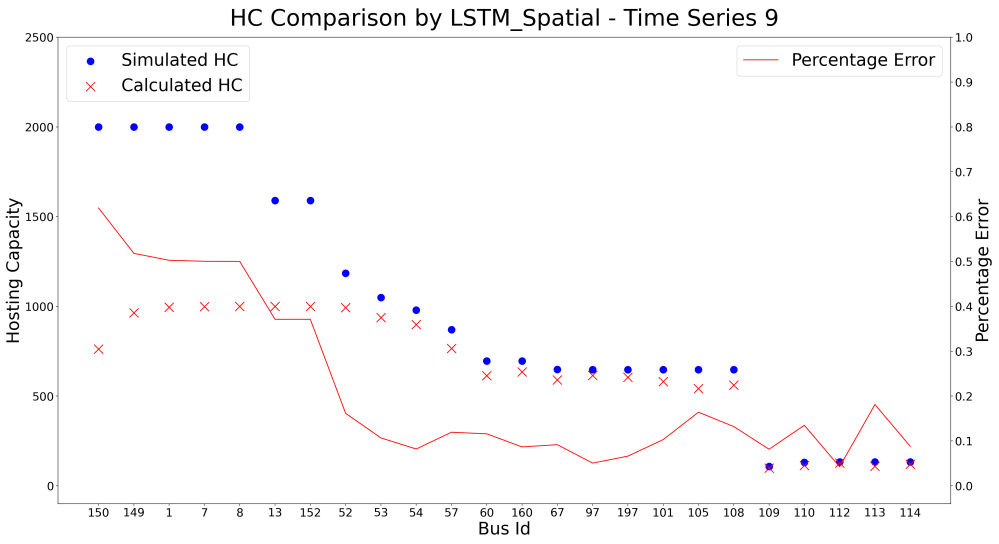
In this section, I will show that the ST-LSTM model is able to build the correlation between spatial sequential and temporal sequential information.

As we can see in Table 6.1, the temporal LSTM has the worst performance. That is because that HC is highly correlated to spatial information while the LSTM only using temporal sequence will ignore this kind of crucial information.

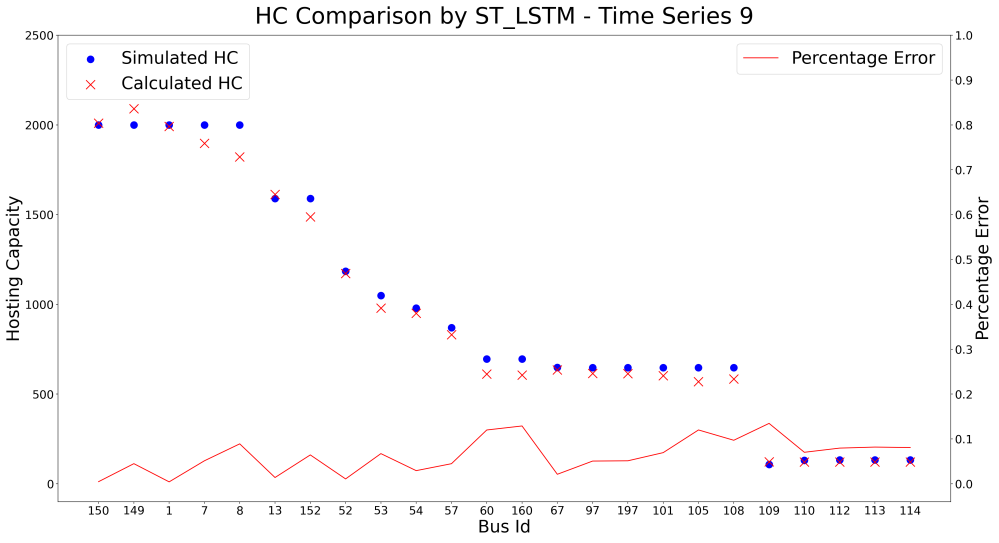
Test Scenario	Mean Square Error	Percentage Error
Temporal LSTM	0.124	252%
Spatial LSTM	0.299	28%
ST-LSTM (without sensitivity gate)	0.0025	5%

Table 6.1: Comparison Between the Basic LSTM and the ST-LSTM on the Longest Path

For the spatial LSTM and the ST-LSTM model, the ST-LSTM model has lower MSE and percentage error. Fig 6.1, 6.2, and 6.3, are the comparisons between these two models at different time-series.

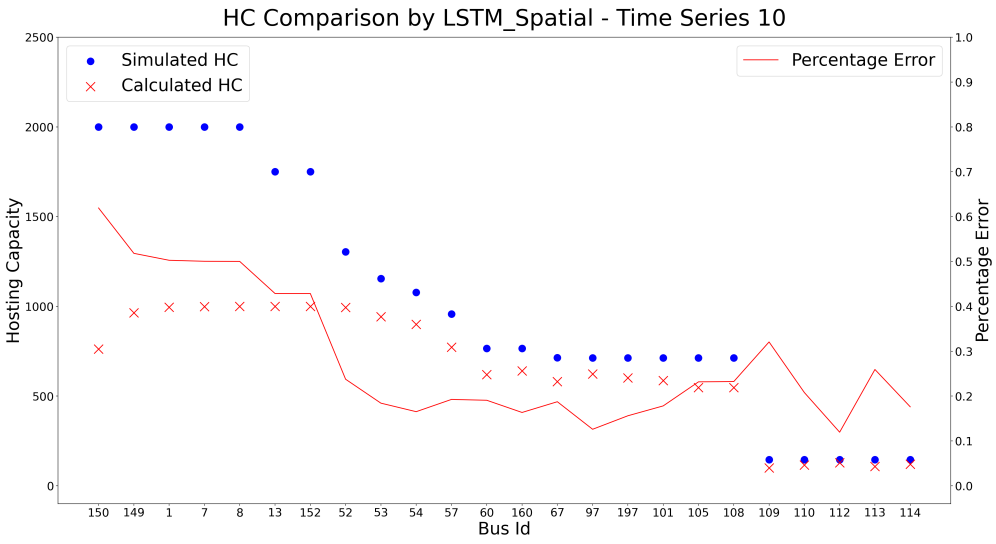


(a) Spatial LSTM Time-series 9

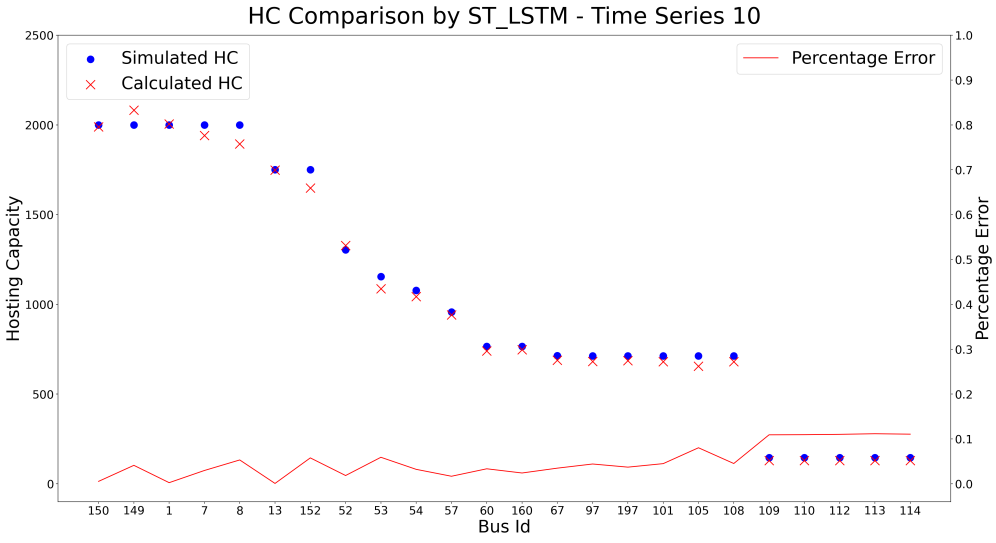


(b) ST-LSTM Time-series 9

Figure 6.1: Result Comparison Between the Spatial LSTM and the ST-LSTM at Time-series 9 on Path 1.

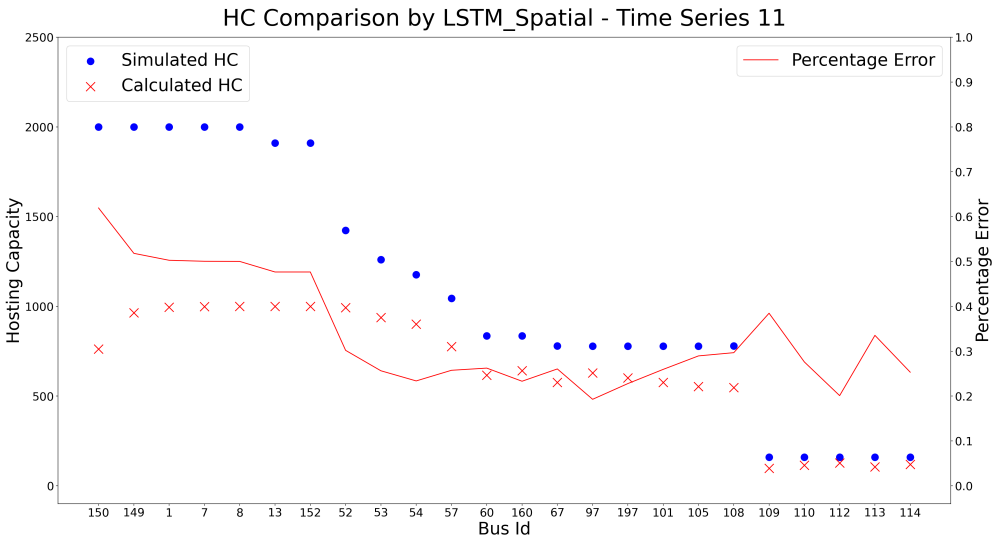


(a) Spatial LSTM Time-series 10

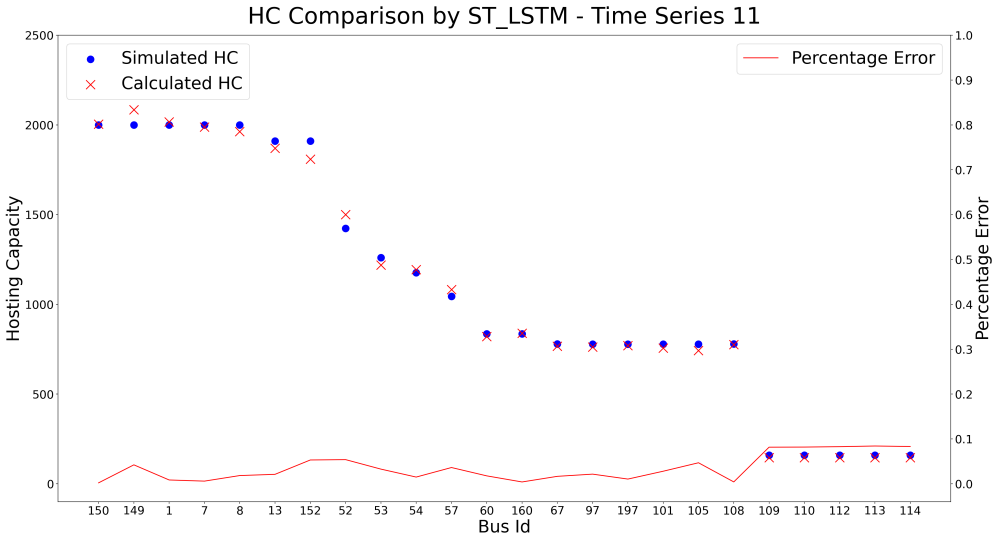


(b) ST-LSTM Time-series 10

Figure 6.2: Result Comparison Between the Spatial LSTM and the ST-LSTM at Time-series 10 on Path 1.



(a) Spatial LSTM Time-series 11



(b) ST-LSTM Time-series 11

Figure 6.3: Result Comparison Between the Spatial LSTM and the ST-LSTM at Time-series 11 on Path 1.

6.4 Results and Comparison on Different Paths

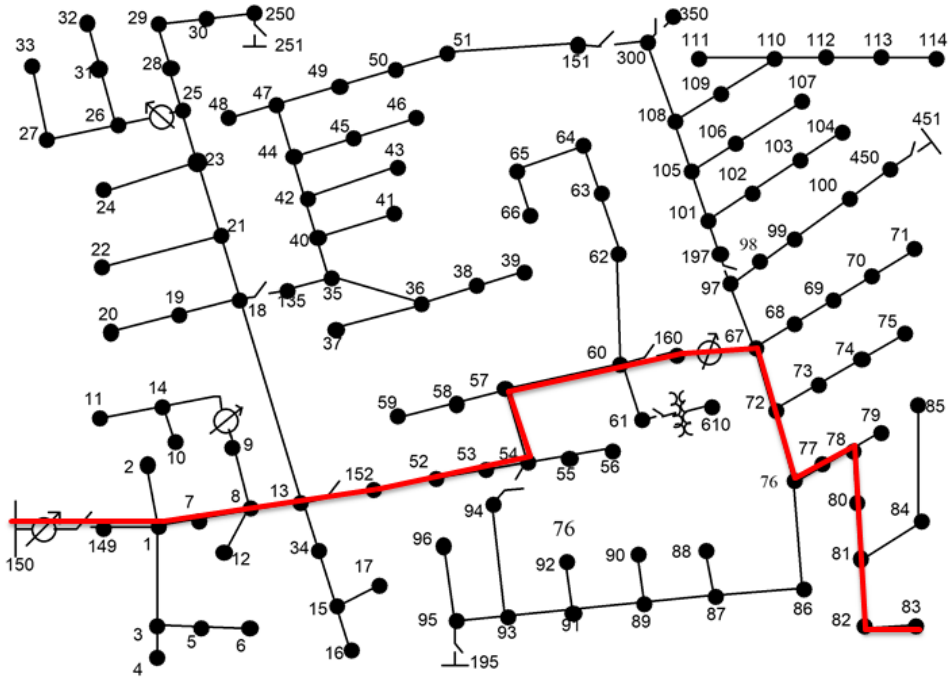
In this section, I will show that ST-LSTM models will significantly increase the accuracy of HC calculation of the whole network.

To illustrate this, I chose another two paths in the IEEE 123-bus system. The numerical result comparison between the spatial LSTM and the ST-LSTM of path 2 and path 3 are shown in Table 6.2. We can see, compared to the spatial LSTM, the ST-LSTM has a significantly smaller percentage error on both of the two paths.

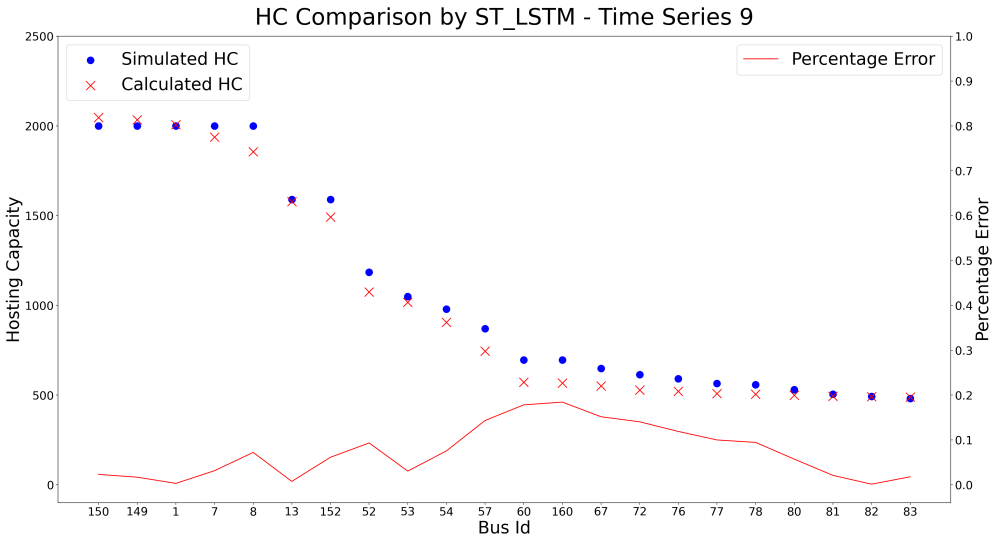
Furthermore, Fig 6.4 shows the detailed testing result of the ST-LSTM on Path 2, and Fig 6.5 shows the detailed testing result of the ST-LSTM on Path 3.

Percentage Error	Path 2	Path 3
Spatial LSTM	28.5%	45.8%
ST-LSTM (without sensitivity gate)	6.5%	6.8%

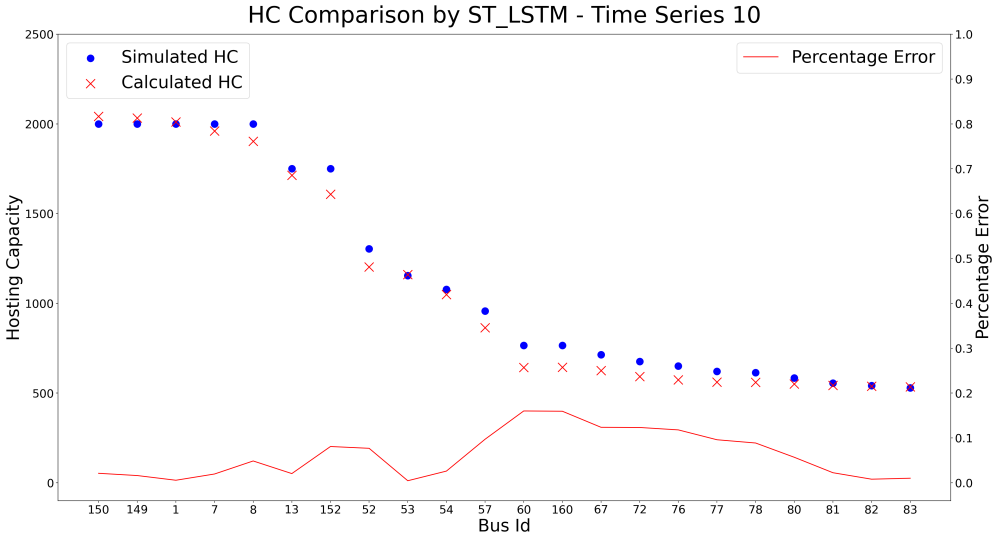
Table 6.2: Percentage Error Comparison Between the Spatial LSTM and the ST-LSTM on Other Paths



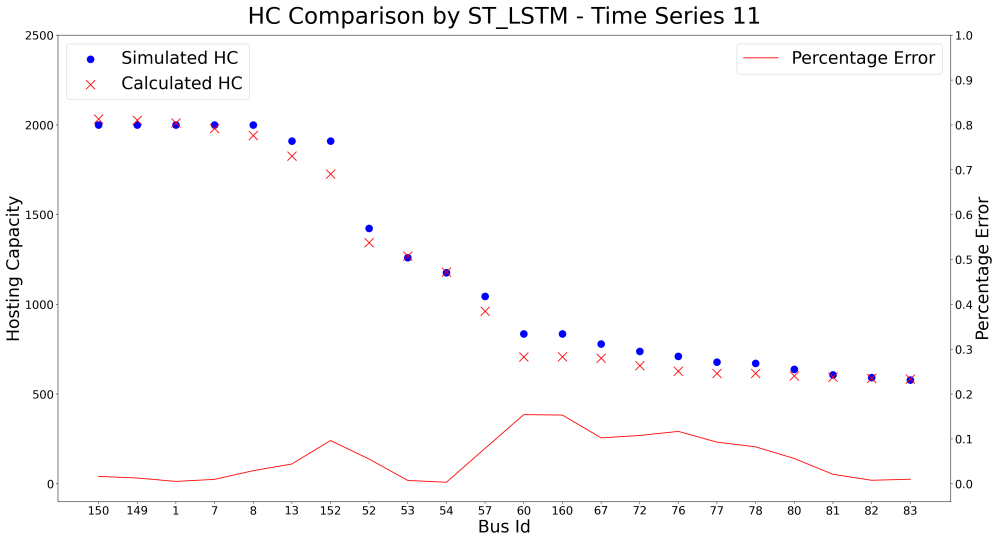
(a) Path 2 in the IEEE 123-bus System



(b) ST-LSTM Time Series 9

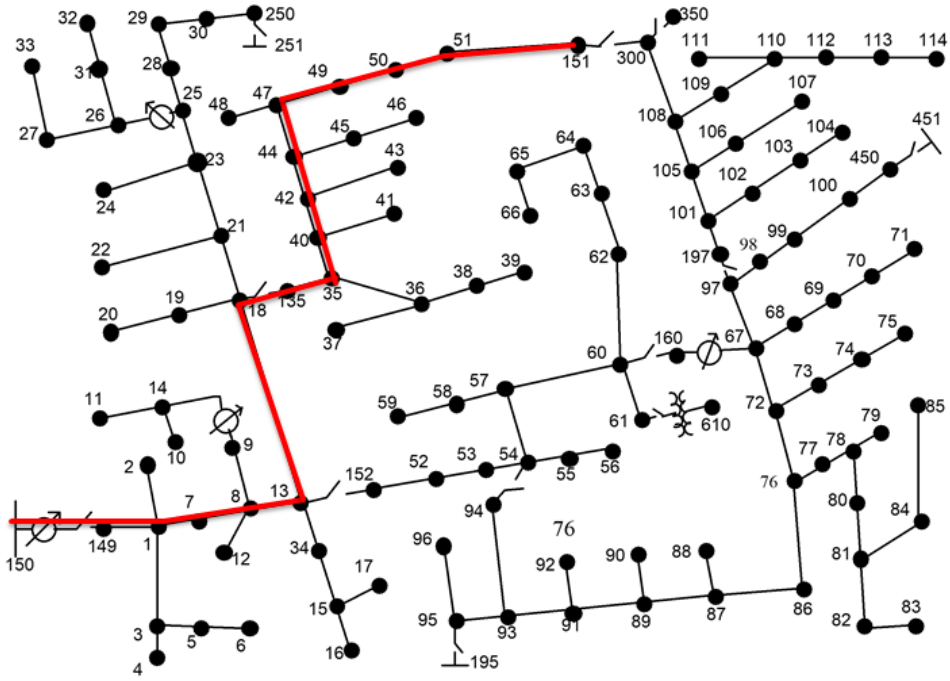


(c) ST-LSTM Time-series 10

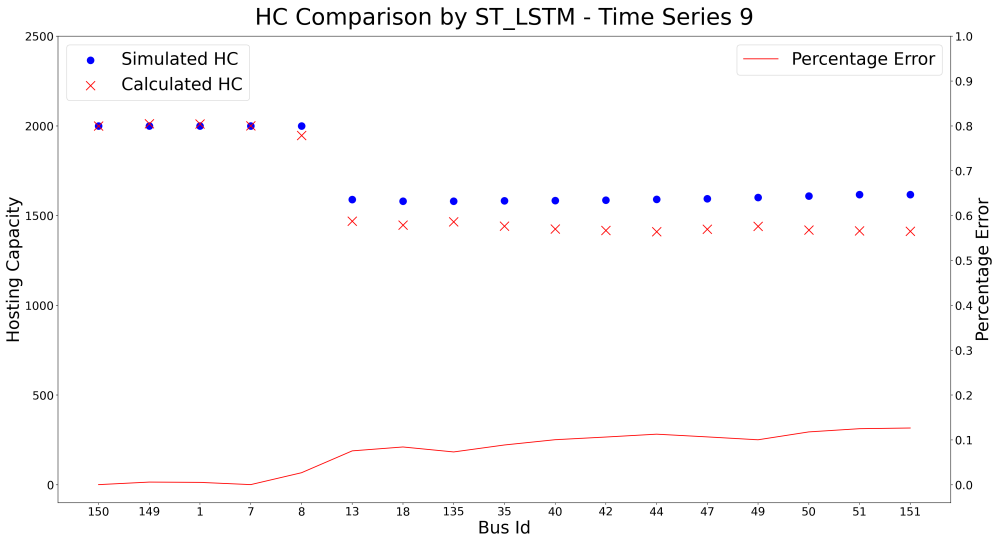


(d) ST-LSTM Time-series 11

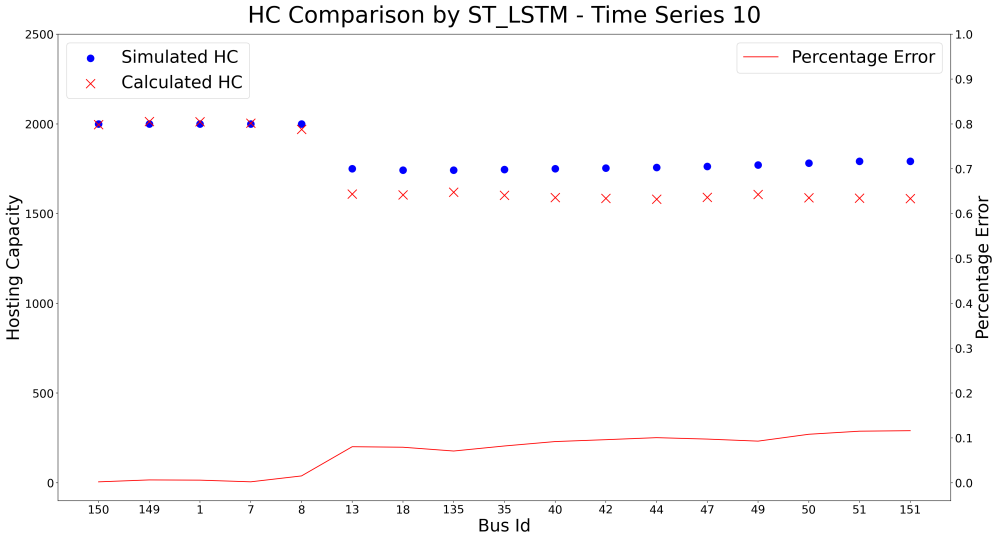
Figure 6.4: Result of the ST-LSTM on Path 2.



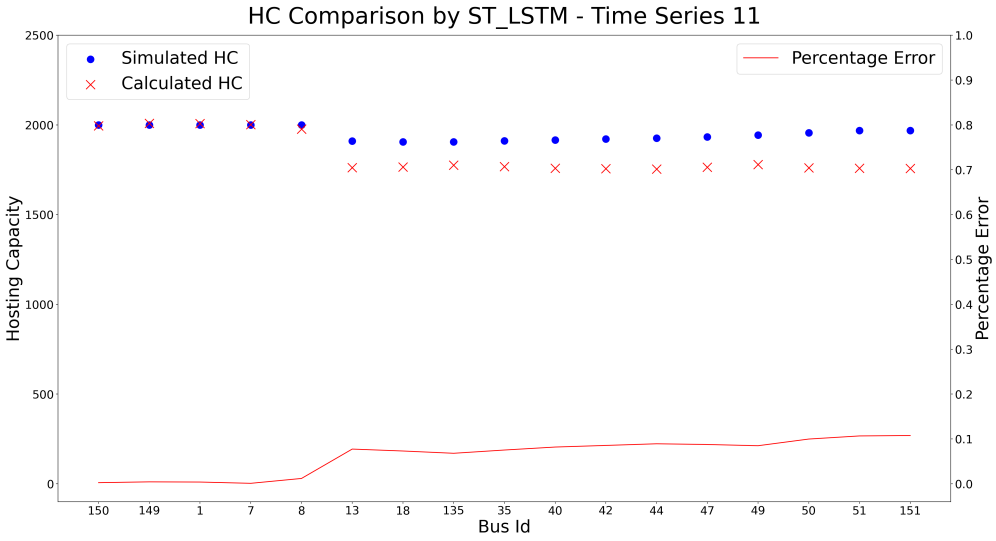
(a) Path 3 in the IEEE 123-bus System



(b) ST-LSTM Time-series 9



(c) ST-LSTM Time-series 10



(d) ST-LSTM Time-series 11

Figure 6.5: Result of the ST-LSTM on Path 3.

6.5 Comparison Between the Spatial LSTM and the ST-LSTM in the Utility Network

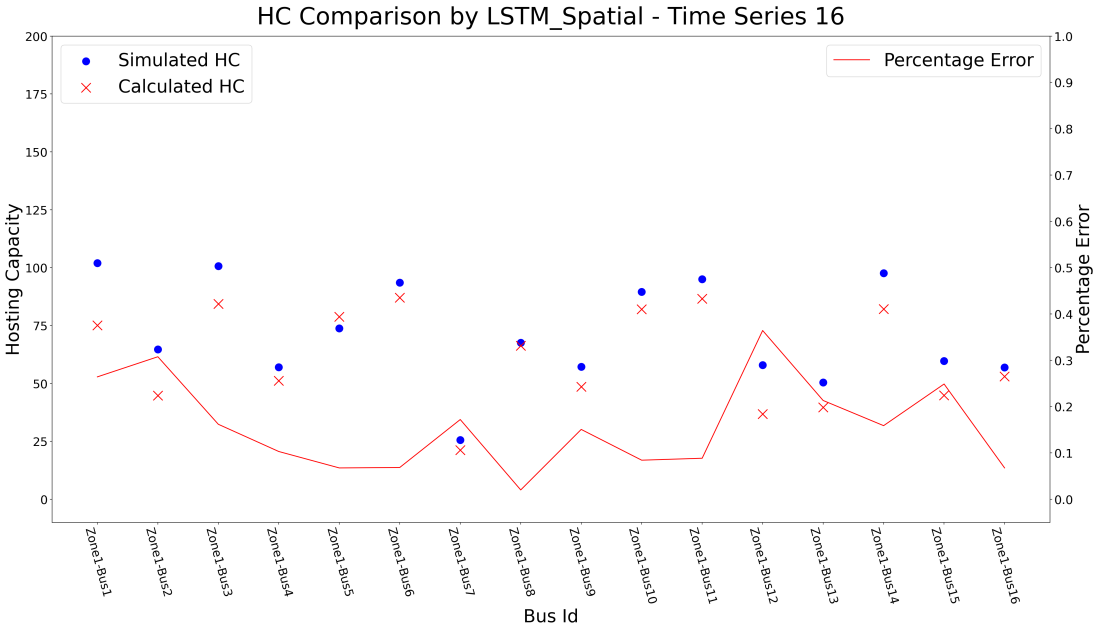
In this section, I will show the result in a real utility network, which will further verify that the ST-LSTM will reduce the complexity in HC analysis because the network is more complicated compared with the IEEE 123-bus network.

I divided the utility network into different zones. Division of the zones was primarily based on the location of the buses and where those buses are connected into the main 3-phase trunk. The comparison is shown in Table 6.3. The percentage error of the ST-LSTM is lower as well compared with the spatial LSTM.

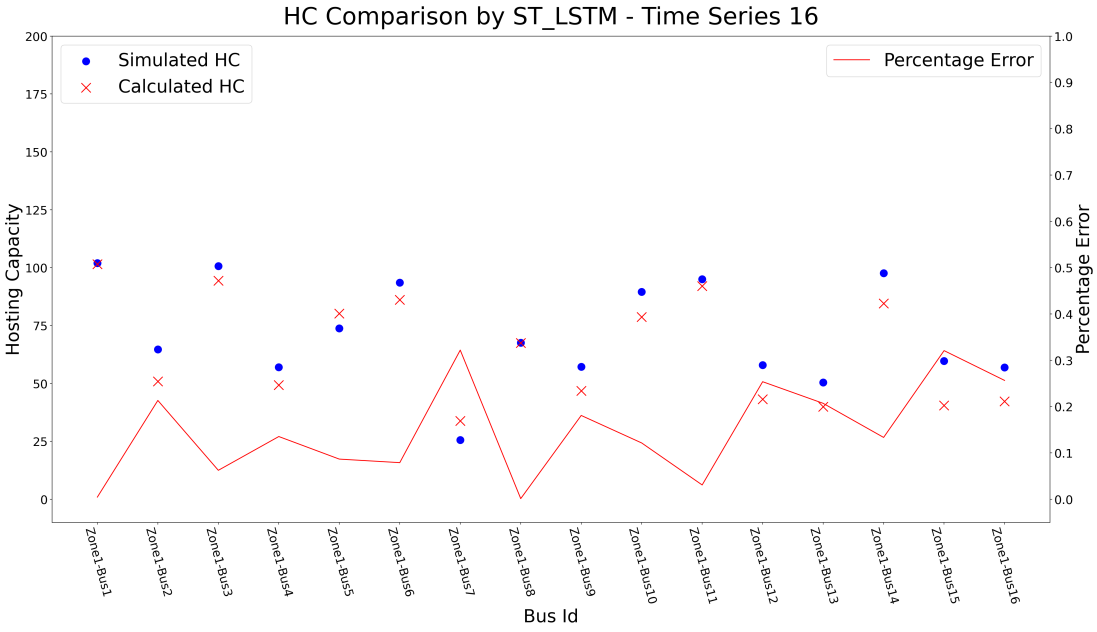
Percentage Error	Zone 1	Zone 2	Zone 3
Spatial LSTM	18.3%	9.3%	12.5%
ST-LSTM (without sensitivity gate)	12.2%	7.2%	12.1%

Table 6.3: Percentage Error Comparison Between the Spatial LSTM and the ST-LSTM in the Utility Network

The result of the zone 1 is shown in Fig 6.6, 6.7, and 6.8.

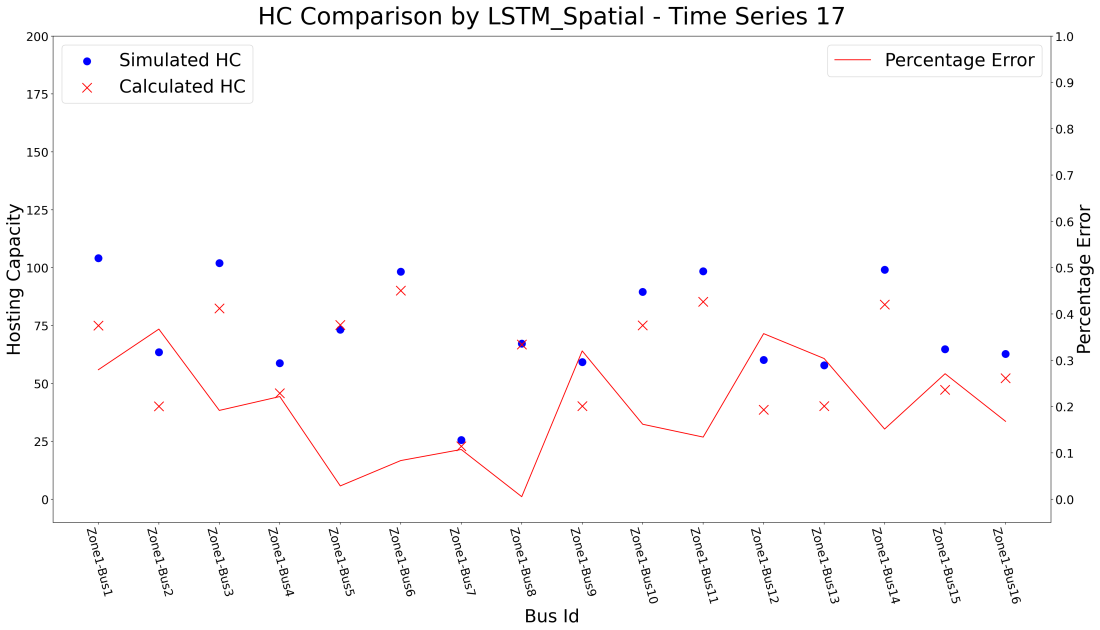


(a) Spatial LSTM Time-series 16

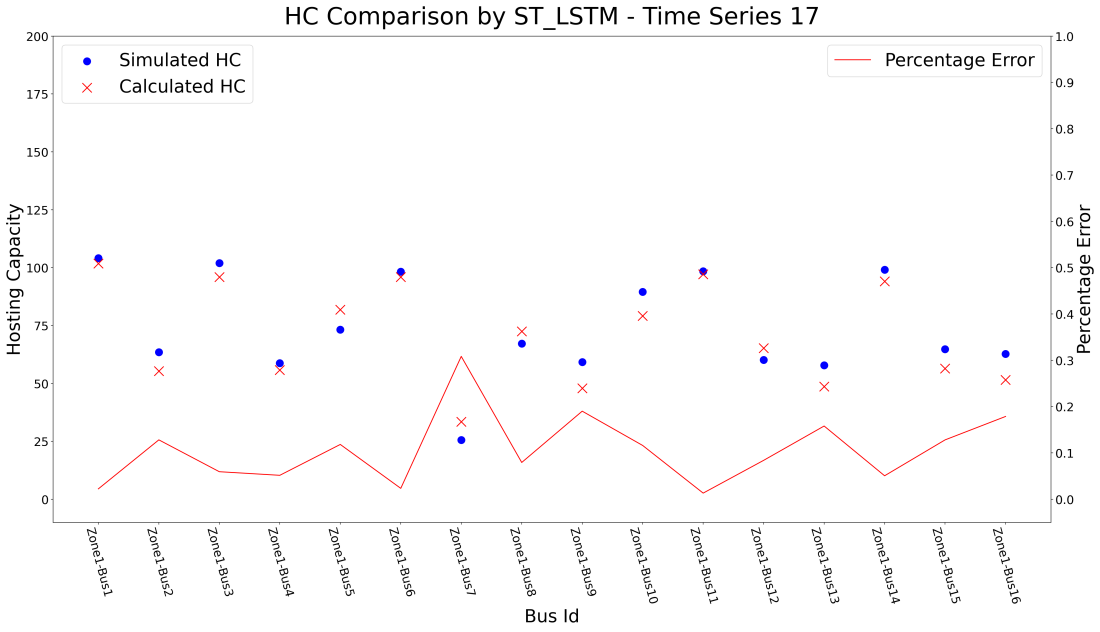


(b) ST-LSTM Time-series 16

Figure 6.6: Result of Time-series 16 in Zone 1 of the Utility Network.

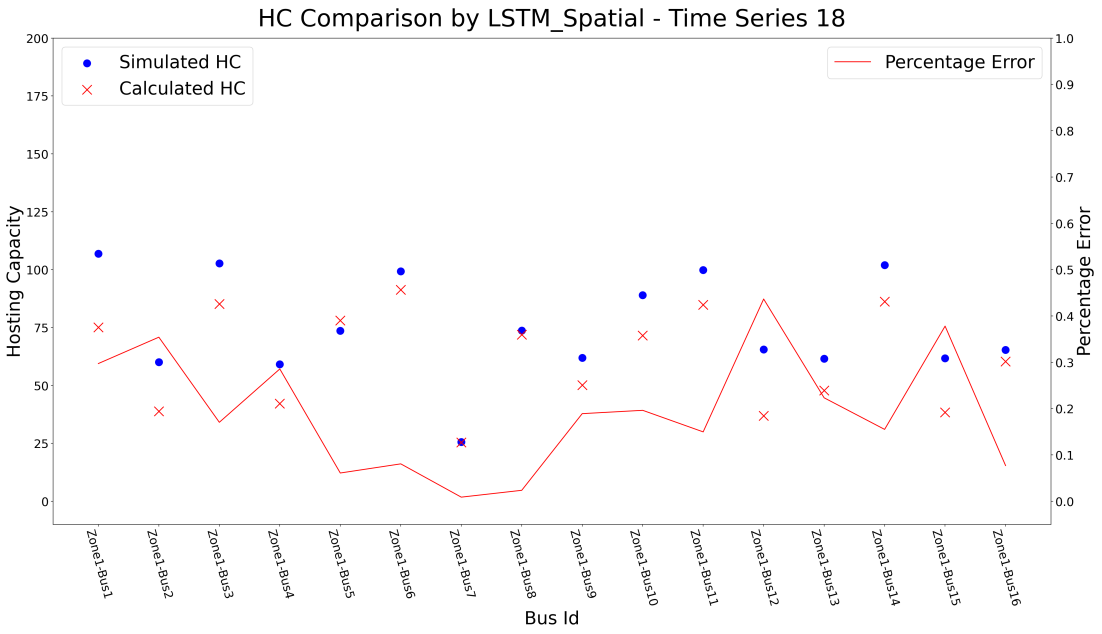


(a) Spatial LSTM Time-series 17

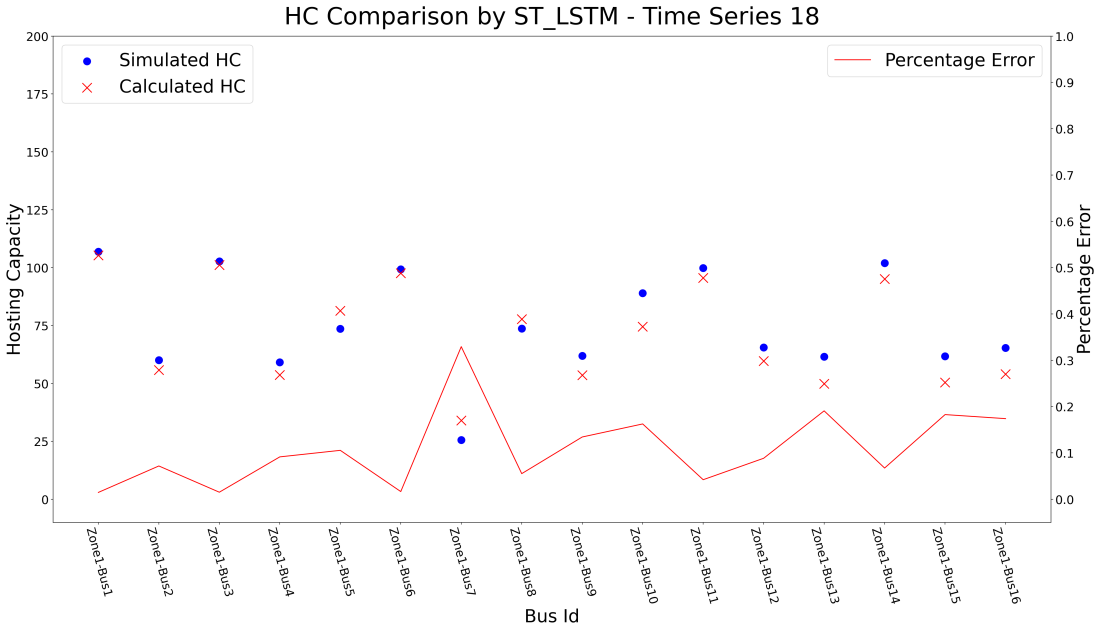


(b) ST-LSTM Time-series 17

Figure 6.7: Result of Time-series 17 in Zone 1 of the Utility Network.



(a) Spatial LSTM Time-series 18



(b) ST-LSTM Time-series 18

Figure 6.8: Result of Time-series 18 in Zone 1 of the Utility Network.

Chapter 7

CONCLUSION

In this paper, I used a modified deep learning-based method, namely Spatial-Temporal LSTM, to calculate the HC in the LV distribution grid. There are three main contributions. First, this method significantly simplifies the HC analysis and calculates HC in real-time. Second, the ST-LSTM builds a correlation between temporal and spatial sequence. Third, this paper increases the accuracy of the deep learning-based HC calculation method by modifying the basic LSTM algorithm.

For the temporal sequence, I used historical power system data. For the spatial sequence, I used the longest path method to construct the sequence.

By conducting the comparison between basic LSTM and the ST-LSTM, I verified that the ST-LSTM has a considerable increase of both the average and the respective HC accuracy at the buses.

To further verify the performance of the ST-LSTM deep learning framework, I used the data from a utility feeder to test the performance. The accuracy is higher as well compared with the basic LSTM.

In the future, I plan to improve the accuracy of the current model by finding more highly relevant input features. Moreover, developing an algorithm to merge all the paths automatically is also listed as my future target.

REFERENCES

[1] 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.

[2] 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.

[3] E. Mulenga, M. H. Bollen, and N. Etherden, "A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids," *International Journal of Electrical Power & Energy Systems*, vol. 115.

[4] J. Smith, R. Dugan, M. Rylander, and T. Key, "Advanced distribution planning tools for high penetration PV deployment," in *IEEE Power and Energy Society General Meeting*, 2012, pp. 1–7.

[5] S. Elsaiah, M. Benidris, and J. Mitra, "Analytical approach for placement and sizing of distributed generation on distribution systems," *IET Generation, Transmission & Distribution*, vol. 8, no. 6, pp. 1039–1049, 2014.

[6] H. Hu, Q. Shi, Z. He, J. He, and S. Gao, "Potential harmonic resonance impacts of PV inverter filters on distribution systems," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 151–161, 2014.

[7] Z. Ren, W. Yan, X. Zhao, Y. Li, and J. Yu, "Probabilistic power flow for distribution networks with photovoltaic generators," in *IEEE Power & Energy Society General Meeting*, 2013, pp. 1–5.

[8] M. H. Bollen and S. K. Rönnberg, "Hosting capacity of the power grid for renewable electricity production and new large consumption equipment," *Energies*, vol. 10, no. 9, p. 1325, 2017.

[9] C. Wan, Z. Xu, Z. Y. Dong, and K. P. Wong, "Probabilistic load flow computation using first-order second-moment method," in *IEEE Power and Energy Society General Meeting*, 2012, pp. 1–6.

[10] J. Quiroz and M. Reno, "Detailed grid integration analysis of distributed PV," in *IEEE Photovoltaic Specialists Conference*, 2012, pp. 596–601.

[11] Y. Xue, I. Sharma, T. Kuruganti, J. Nutaro, J. Dong, M. Olama, and D. Fugate, "Voltage impact analyses of solar photovoltaics on distribution load tap changer operations," in *North American Power Symposium*, 2017, pp. 1–6.

[12] M. Al-Saffar and P. Musilek, "Reinforcement learning-based distributed bess management for mitigating overvoltage issues in systems with high PV penetration," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2980–2994, 2020.

[13] Y. Liu, N. Zhang, Y. Wang, J. Yang, and C. Kang, "Data-driven power flow linearization: A regression approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2569–2580, 2018.

[14] S. Breker, J. Rentmeister, B. Sick, and M. Braun, "Hosting capacity of low-voltage grids for distributed generation: Classification by means of machine learning techniques," *Applied Soft Computing*, vol. 70, pp. 195–207, 2018.

[15] X. Geng, L. Tong, A. Bhattacharya, B. Mallick, and L. Xie, "Probabilistic hosting capacity analysis via bayesian optimization," *arXiv preprint arXiv:2011.05193*, 2020.

[16] P. Zhao, H. Zhu, Y. Liu, Z. Li, J. Xu, and V. S. Sheng, "Where to go next: A spatio-temporal LSTM model for next poi recommendation," *arXiv preprint arXiv:1806.06671*, 2018.

[17] L. Huang, Y. Ma, S. Wang, and Y. Liu, "An attention-based spatiotemporal lstm network for next poi recommendation," *IEEE Transactions on Services Computing*, 2019.

[18] D. Kong and F. Wu, "HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction." in *International Joint Conferences on Artificial Intelligence*, vol. 18, no. 7, 2018, pp. 2341–2347.

[19] 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.

[20] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.

[21] R. A. Shayani and M. A. G. de Oliveira, "Photovoltaic generation penetration limits in radial distribution systems," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1625–1631, 2010.

[22] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*, 2013, pp. 1310–1318.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.