A spatial-temporal ensemble deep learning framework for wind power forecasting (Team QDU)

Zhiruo Li
Qingdao University
Qingdao, Shandong, China

Jieqi Xing
Qingdao University
Qingdao, Shandong, China

Shunyao Wu*
Qingdao University
Qingdao, Shandong, China
wushunyao@qdu.edu.cn

ABSTRACT

Wind Power is a promising renewable source which has become widely used in recent years. Wind power forecasting is a critical issue which is beneficial to maintain the balance between power generation and consumption. However, the problem is also challenging due to limited information and the variety of time series. In this paper, we proposed an efficient deep learning framework to address the spatial dynamic wind power forecasting challenge in Baidu KDD Cup 2022. In our solution, we constructed three kinds of deep learning models utilizing both time series and spatial distribution and finally integrated them by model averaging. Our method is highly efficient and robust which can achieve a score of -46.18 in the phase 3. The source code is available at https://github.com/hansu1017/SDWPF-Baidu-KDD-Cup-2022.

KEYWORDS

Deep learning, Spatial distribution, Wind power forecasting

ACM Reference Format:
Zhiruo Li, Jieqi Xing, and Shunyao Wu. 2018. A spatial-temporal ensemble deep learning framework for wind power forecasting (Team QDU). In *Proceedings of (Baidu KDD Cup 2022, Mar. 16 – Jul. 17, 2022).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Wind energy is a clean and safe renewable resource which plays an important role for a sustainable future. However, the power generated by wind turbines varies rapidly due to the fluctuation of wind speed and wind direction as well as the terrain, humidity, date and time of the day[6]. Thus, a balance is highly desired between the power supply and demand[4]. It's necessary to explore an accurate wind power forecasting method which can effectively reduce the enormous impact on grid operation safety when high permeability intermittent power supply is connected to the power grid[1]. Multiple methods have been proposed to address this problem such as physical methods[2], statistical methods[5] and deep learning models[3].

The Baidu KDD Cup proposed a spatial dynamic wind power forecasting (SDWPF) challenge which aims to predict the active power of the next 288 time points for each turbine providing participants with the dynamic context and spatial distribution[7]. However, the SDWPF problem is still challenging. First, how to effectively integrate spatial-temporal information is critical. Second, the characteristic of multiple time series is difficult to be captured, which makes it challenging to precisely forecast the relatively long-term active power sequence. Third, the original dataset contains a large amount of unknown values, hence how to process the data properly is also challenging.

To address the above challenges, we proposed a spatial-temporal ensemble deep learning framework which can obtain a relatively accurate and robust prediction for SDWPF. To effectively integrate spatial-temporal information, we utilized both time series and space distribution information as features to comprehensively forecast the wind power. Besides, we developed three kinds of deep learning models to fully exploit the spatial-temporal information. In addition, to improve the robustness of our method, we sampled the data with different seeds to enrich the training set and dropped samples with too many unknown values. Our proposed framework achieved an score of -46.18 in the phase 3 which ranked 13 in the paddlepaddle track.

2 PROPOSED FRAMEWORK

Figure 1 illustrates the main flow of our proposed framework. Firstly, five variables including Wspd (Wind speed), Pab1 (Pitch angle of blade 1), Etmp (Temperature of the surrounding environment), Itmp (Temperature inside the turbine nacelle), Patv (Active power) along with the spatial distribution information were selected from all the available information according to our multiple attempts. Secondly, we generated sequence features from the five variables and obtained three small datasets by random sampling and outlier elimination. Then, three deep learning models based on CNN and GRU were constructed to integrate the spatial-temporal information. Finally, the forecast values were assembled by model averaging to improve the robustness.

2.1 Preprocessing

The SDWPF dataset provided by Baidu contains a wind power dataset and a location dataset. The wind power dataset sampled data from every 10 minutes from each wind turbine in the farm and the total days and turbines are 245 and 134 respectively. The location dataset provided the relative position of all turbines[7]. To construct training set, for the turbine i at timestamp t, we took the time series from $t-144$ to t as features and the Patv from $t+1$ to $t+288$ as the target to be predicted. Due to the large sample size of the training

---
*Corresponding Author.

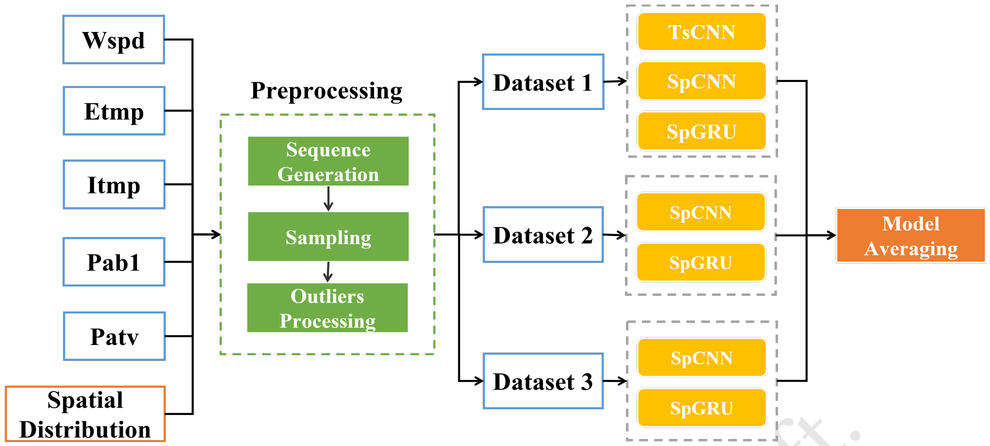

Figure 1: Our proposed framework for the spatial dynamic wind power forecasting

set and the large number of sequences with overlap, the training set of each turbine was randomly sampled by a ratio of 0.01 and the sampled data of all turbines were subsequently concatenated and shuffled, which not only performed better in offline experiments, but also remarkably improved the training speed. To fully make use of the temporal information, We generated three datasets by repeating the above steps with three different seeds which enhanced the robustness of models.

Besides, we also processed the samples in the training set with too many unknown values. According to the official report[7], samples with Patv less than 0 or Pab1 more than 89° can be considered as unknown values. By data analysis, we found a considerable percentage of unknown samples in the training sets, which had significantly negative impact on the accuracy of SDWPF. Thus, we dropped the samples with at least one sequence that contains more than 100 unknown values.

2.2 Feature engineering

Through offline experiments, we found that using all sequences is not as good as using part of sequences including Wspd sequence, Patv sequence, Pab1 sequence, Etmp sequence and Itmp sequence. For the two sequences related to temperature, we constructed the absolute difference between Etmp and Itmp as a new sequence which was denoted as Temp. Besides, for all models, we introduced a spatial Patv sequence based on space distribution. More specifically, for each turbine of each timestamp, we calculated the Euclidean distance with other turbines. Then, we selected the 121 closest turbines and used their Patvs to construct a new sequence denoted as spatial-Patv.

2.3 Deep learning models with spatial-temporal data

We trained three kinds of deep learning models with different structures including a CNN model with multichannel time series (TsCNN), a CNN model with spatial-temporal information (SpCNN) and a GRU model with spatial-temporal information (SpGRU). The first model only trained by one dataset, while the others were trained using three datasets to fully make use of the temporal-spatial information. Additionally, the final predictions were obtained by calculating the arithmetic mean of all model outputs.

2.3.1 *CNN model with multichannel time series.*

Figure2(a) demonstrates the structure of the model. By data analysis, we found the time series had a cyclical pattern, thus we decided to use convolution kernels to learn the cyclical pattern. Specifically, we transformed the four sequences including Wspd sequence, Patv sequence, Pab1 sequence and the Tmp sequence into a 4-dimensional array with size 12*12*4. Then, the map was input to a CNN layer with 3*3 convolution kernels and a residual network which is also a CNN with 1*1 kernels. Then, the feature maps of the two network were added and continuously input to two similar networks. Finally, the output went through a third CNN layer and obtain the predictions by a fully connected layer. Each CNN layer contains an average pooling which equivalent to do moving average for time series in order to further learn the cyclical pattern.

2.3.2 *CNN model with spatial-temporal information.*

Unlike the CNN based model above, in this model, the length of each time series sequence was 121. Besides, we introduced the spatial feature to the CNN model. With regard to the model structure, as shown in Figure2(b), for time series, we simplified the CNN layers with multichannel time series to two and added a CNN layer to learn spatial patterns. It was noteworthy that in the CNN layer 3, we replaced the mean-pooling by max-pooling to highlight the neighbor turbines with strong influence.

2.3.3 *GRU model with spatial-temporal information.*

As shown in Figure2(c), in this model, four 144-length sequences including Wspd sequence, Patv sequence, Itmp sequence and Etmp sequence were extended to 288-length by complementary zeros, which were subsequently input to a GRU layer to learn the trend of time series. Meanwhile, the spatial feature went through a CNN layer to learn the spatial distribution. Finally, the output of two kinds of layers were concatenated and transformed to the final prediction by a fully connected layer.

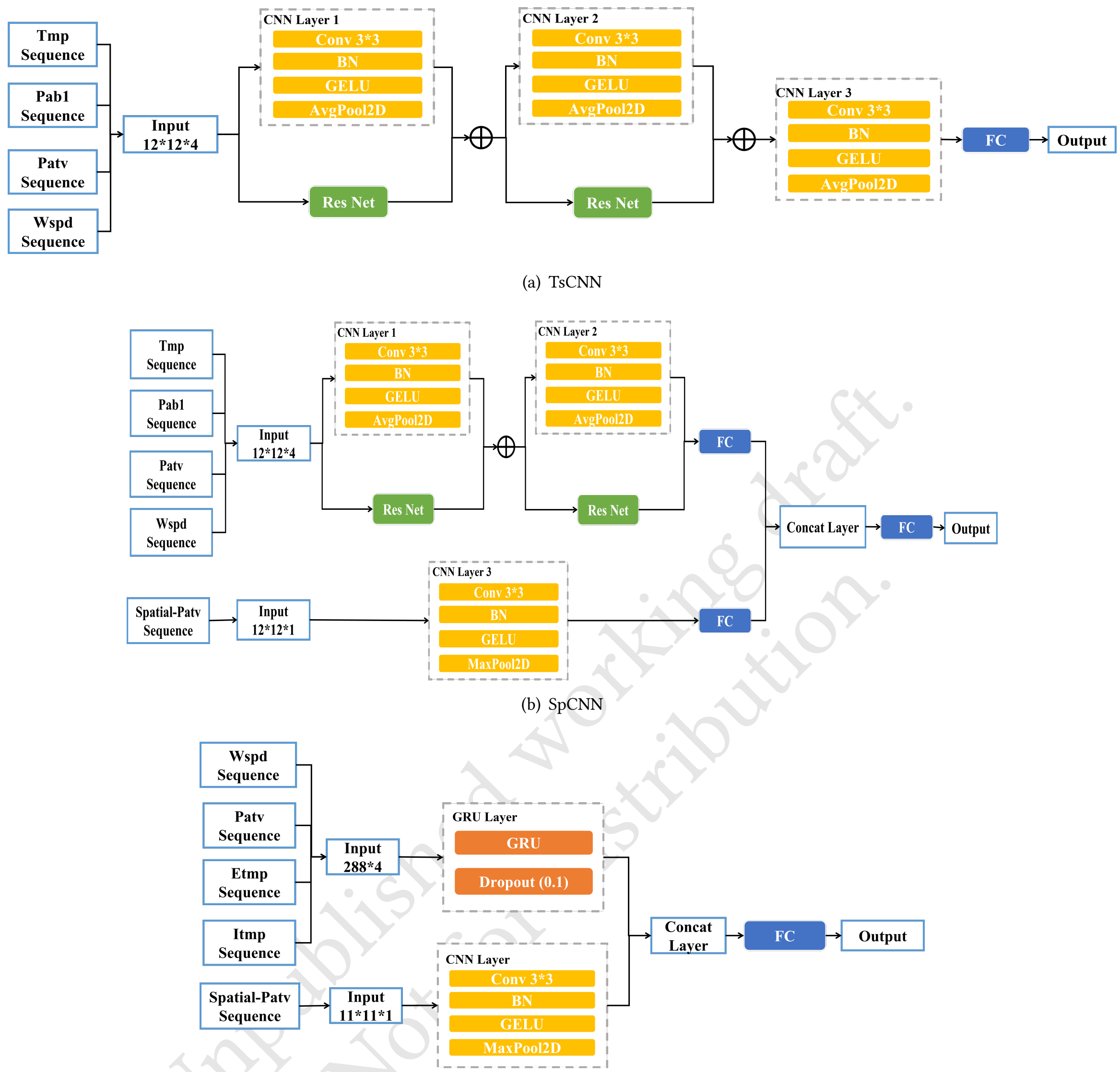

Figure 2: Deep learning models in proposed framework. (a) CNN model with multichannel time series (TsCNN). (b) CNN model with spatial-temporal information (SpCNN). (c) GRU model with spatial-temporal information (SpGRU).

3 EXPERIMENTS

3.1 Offline evaluation

For offline evaluation, we not only used the official file 0001*out.csv* as the validation set (Offline1), but also constructed our own offline validation sets (Offline2). Specifically, from the day 226, the data of every two days thereafter were treated as a validation set and the total number of our own validation files is 30. Additionally, since we predicted the Patv of all turbines by using the same models, we optimized the offline evaluation process which calculated the score of all turbines at once to significantly reduce the evaluation time.

3.2 Performance

Figure 3 displays the predicted curves of the three models using dataset1 as training set and 0001*out.csv* as validation set. As can be seen in the figure, the predictions of the three models are close but still exist differences. The CNN model with multichannel time series is good at learning large fluctuations, while the CNN with spatial-temporal information tends to learn small fluctuations better. In addition, the GRU model with spatial-temporal information has a smoother predicted curve and can better forecast the overall trend of the next 288 time points.

Table 1 illustrates the offline and online results of different frameworks. It can be found that using one kind of model has inconsistent

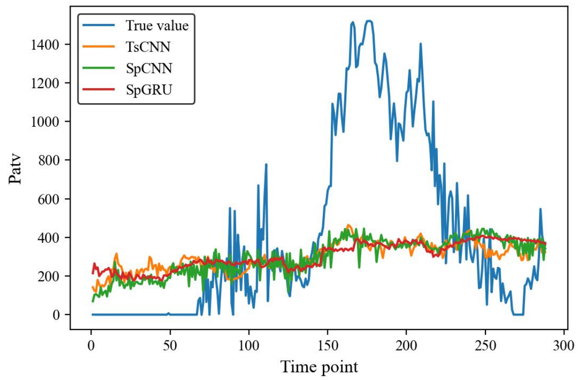

Figure 3: Predicted curves of three models in our proposed framework

Table 1: Performance comparison

Method	Offline1	Offline2	Online
TsCNN	-42.44	-46.27	-
3 SpCNNs	-44.66	-44.35	-
3 SpGRUs	-45.14	-42.72	-
TsCNN + 3 SpCNNs	-43.94	-44.63	-46.30
Proposed framework	-44.36	-43.36	-46.18

performance on different validation sets, while model averaging can apparently improve the robustness of the final performance. Our proposed framework outperformed other methods on both Offline2 and online validation sets, which can achieve a score of -43.36 and -46.18 respectively. In addition, our proposed framework is relatively efficient since the evaluation time for 30 predictions is 194.76 seconds on a Linux machine with Tesla V100 GPU.

4 CONCLUSIONS

In this paper, our team proposed an effective deep learning framework to address the spatial dynamic wind power forecasting challenge. To enhance the efficiency and robustness of the models, we generated and sampled the sequences and dropped the samples with too many unknown values. By performing offline experiments, we selected five sequences and constructed two new sequences including Tmp sequence and spatial-Patv sequence. As for models, we developed three kinds of deep learning models including CNN model with multichannel time series, CNN model with spatial-temporal information and GRU model with spatial-temporal information and integrated their predictions by model averaging.

5 ACKNOWLEDGMENTS

This paper is supported by the grant 21BTJ045 from National Social Science Found of China. We thank everyone associated with organizing and sponsoring the Baidu KDD Cup 2022. Dataset was provided by Baidu and the challenge was sponsored and managed by the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. We are very grateful to the KDD Cup Chairs and the staff of Baidu for their great efforts during the challenge.

REFERENCES

[1] Xing Deng, Haijian Shao, Chunlong Hu, Dengbiao Jiang, and Yingtao Jiang. 2020. Wind power forecasting methods based on deep learning: A survey. *Computer Modeling in Engineering and Sciences* 122, 1 (2020), 273.

[2] Ulrich Focken, Matthias Lange, and Hans-Peter Waldl. 2001. Previento-a wind power prediction system with an innovative upscaling algorithm. In *Proceedings of the European Wind Energy Conference, Copenhagen, Denmark*, Vol. 276. Citeseer.

[3] GN Kariniotakis, GS Stavrakakis, and EF Nogaret. 1996. Wind power forecasting using advanced neural networks models. *IEEE transactions on Energy conversion* 11, 4 (1996), 762–767.

[4] Rahul Sharma and Diksha Singh. 2018. A review of wind power and wind speed forecasting. *Journal of Engineering Research and Application* 8, 7 (2018), 1–9.

[5] George Sideratos and Nikos D Hatziargyriou. 2007. An advanced statistical method for wind power forecasting. *IEEE Transactions on power systems* 22, 1 (2007), 258–265.

[6] Shikha Singh, TS Bhatti, and DP Kothari. 2007. Wind power estimation using artificial neural network. *Journal of Energy Engineering* 133, 1 (2007), 46–52.

[7] Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Yu Li, Ji Liu, Jiantao Su, Junfu Lyu, Yanjun Ma, and Dejing Dou. 2022. SDWPF: A Dataset for Spatial Dynamic Wind Power Forecasting Challenge at KDD Cup 2022. (2022).