Temporal Transfer Learning for Ozone Prediction based on CNN-LSTM Model

Tuo Deng[1], Astrid Manders[2], Arjo Segers[2], Yanqin Bai[3] and Hai Xiang Lin[1]
[1]*Delft Institute of Applied Mathematics, Delft University of technology, Mekelweg 4, 2628 CD Delft, The Netherlands*
[2]*TNO, Climate Air and Sustainability, Utrecht, 3584 CB, The Netherlands*
[3]*Department of Mathematics, Shanghai University, Shanghai 200444, China*

Keywords: Short-term Ozone Prediction, Transfer Learning.

Abstract: Tropospheric ozone is a secondary pollutant which can affect human health and plant growth. In this paper, we investigated transferred convolutional neural network long short-term memory (TL-CNN-LSTM) model to predict ozone concentration. Hourly CNN-LSTM model is used to extract features and predict ozone for next hour, which is superior to commonly used models in previous studies. In the daily ozone prediction model, prediction over a large time-scale requires more data, however, only limited data are available, which causes the CNN-LSTM model to fail to accurately predict. Network-based transfer learning methods based on hourly models can obtain information from smaller temporal resolution. It can reduce prediction errors and shorten run time for model training. However, for extreme cases where the amount of data is severely insufficient, transfer learning based on smaller time scale cannot improve model prediction accuracy.

1 INTRODUCTION

Tropospheric ozone is formed by the chemical reaction of precursors rather than directly emitted, so it is classified as a secondary pollutant (McKee, 1993). Studies have shown that tropospheric ozone will seriously affect human health (Council et al., 1992) and plant growth (Iglesias et al., 2006). Therefore, it is very important to make short-term forecasts for ozone concentration in the troposphere.

However, it is very difficult to accurately predict ozone even in the short term. The formation of ozone is determined by complex chemical reactions. Meanwhile, ozone concentration is easily affected by ozone precursor emissions (Placet et al., 2000). Common ozone precursors include nitrogen oxides (NOx) and volatile organic compounds (VOCs), etc. Various industrial activities, vehicular traffic, and agricultural activities generate a large amount of ozone precursors, which can make large changes in ozone concentration in a short time. At the same time, the ozone concentration is also affected by many meteorological factors, such as temperature, wind direction, wind speed, humidity, solar radiation, etc (Council et al., 1992). The day-to-day variability in meteorology and precursor emissions makes it difficult to predict ozone concentrations.

Currently, there are two main approaches to predict ozone concentration. The first approach is based on Chemical Transport Models (CTM). For instance, Lotos-Euros chemical transport model was used to simulate ozone concentration over Europe (Curier et al., 2012). The agreement (temporal correlation coefficient) between in-situ and modeled ozone concentration is good. However, it is difficult for the model to capture the ozone peak during the experiment and it turns to underestimates the daily ozone maximum. Besides, CTMs also have limitations. It implies that some other important errors such as the errors in atmospheric chemistry mechanism can not be ignored (Tang et al., 2011). In addition, the resolution is limited so it cannot resolve all local factors.

At the same time, CTMs often need to simulate complex physical and chemical processes, which consume a lot of time in calculations. Therefore, some researchers use another approach to predict ozone. The second approach is to ignore the complex chemical process in the formation of ozone and directly predict ozone through data-based machine learning methods. For instance, regression tree (Zhan et al., 2018) and Multilayer Perceptron (MLP) (Feng et al., 2019) are studied for short-term ozone forecast.

Long short-term memory (LSTM) is a widely used machine learning model (Hochreiter and

Deng, T., Manders, A., Segers, A., Bai, Y. and Lin, H.
Temporal Transfer Learning for Ozone Prediction based on CNN-LSTM Model.
DOI: 10.5220/0010301710051012
In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) - Volume 2*, pages 1005-1012
ISBN: 978-989-758-484-8

Schmidhuber, 1997) which can be used in time series problems. It can capture long-term dependencies in time-series forecast problems, such as ozone prediction (Eslami et al., 2019). At the same time, Convolutional Neural Network (CNN) is also used by some researchers in air pollution prediction due to its excellent performance in feature extraction (Sayeed et al., 2020). It can extract important features in the input data and improve the performance of model. Many research also combine CNN with LSTM to extract temporal and spatial features. CNN-LSTM has been used in many fields, such as natural language processing (Wang et al., 2016), medical field (Oh et al., 2018) and industrial area (Zhao et al., 2017).

However, data-based machine learning models require large amounts of data to train the model. For large temporal resolution air pollution data sets, it is often difficult to obtain sufficient data. Transfer learning (TL) can be a good way to solve this problem. TL is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem (Tan et al., 2018). There are few studies in air pollution using transfer learning. Ma (Ma et al., 2019) used transfer learning to predict particulate matter on different time scales in China. Compared with particulate matter, ozone has a more obvious periodicity, which also provides a basis for our experiments.

This paper aims to use CNN-LSTM to predict ozone concentration and use Transfer Learning to fit the new model for larger temporal resolution. We first use CNN-LSTM to fit hourly ozone prediction model. Compared to other commonly used models in previous studies, the CNN-LSTM model has smaller prediction errors and it can predict the trend of ozone well. For daily ozone prediction model, we do not have sufficient data from daily data set to get accurate result from CNN-LSTM model. With the hourly model as a basic model, we use transfer learning to get new daily ozone prediction model, i.e., TL-CNN-LSTM model. Compared with the LSTM and CNN-LSTM model, our TL-CNN-LSTM has significantly improved the prediction in terms of root mean square error, Pearson correlation coefficient and run time for model training. It implies that transfer learning can obtain knowledge from a smaller temporal resolution model and improve model prediction accuracy.

2 DATA AND ANALYSIS

In this section, we describe the processing and analysis of data. We use interpolation to fix the missing data and merge the two data sets on different time scales. Then we conduct correlation analysis for ozone concentration. For the particularity of time series, we also need to reconstruct the input data.

2.1 Data Description

The air pollution data that we used is provided by German environmental agency (UBA). Specifically, we use the air pollution data of Eisenhüttenstadt station 'DEBB032' from 2014 to 2018, which contains 43824 hourly observations. This site was selected in the experiment because the site has less missing data, which can reduce the error caused by interpolation. Each set corresponds to hourly measurements, including CO, NO_2, NO, NO_X, O_3, PM_{10}, $PM_{2.5}$ and SO_2. The description of station DEBB032 is in Table 1. To understand the distribution of ozone, we plot a bar chart with ozone concentration. We use intervals of 10 $\mu g/m^3$ and show the distribution of ozone in Figure 1. We can see that ozone is concentrated in the range of 40 to 80 $\mu g/m^3$. There are only a few cases where the ozone concentration is higher than 100 $\mu g/m^3$.

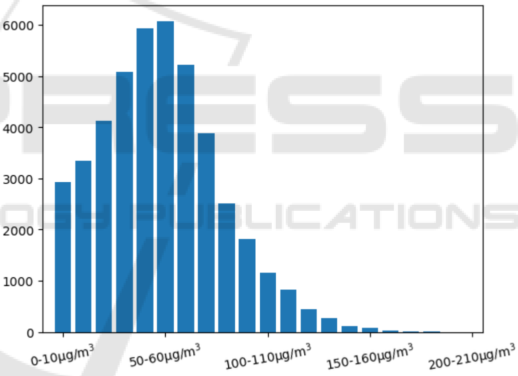


Figure 1: Distribution of ozone concentration.

In order to simulate the impact of meteorology on ozone concentration, we also selected the corresponding meteorological data in the E-OBS data set as a supplement (https://www.ecad.eu/). E-OBS data set is an ensemble dataset based on gridded observation of national meteorological institutes and available on a 0.1 and 0.25 degree regular grid. It covers the area: 25N-71.5N x 25W-45E. In the experiment, five meteorological features were selected, including the elements daily mean temperature TG, daily minimum temperature TN, daily maximum temperature TX, daily precipitation sum RR and daily averaged sea level pressure PP. In 0.25°*0.25° data set, we select the meteorological data of the grid point closest to the station (52.125°N,12.625°E) at the corresponding time.

Table 1: Station Description.

No. of station	Lat	Lon	Airbase station type	Airbase station ozone classification
DEBB032	52.146264	14.638166	industrial	suburban

2.2 Data Interpolation and Merging

Due to instrument malfunction and other reasons, there are missing values in the air pollution data set. The missing rate of each features is less than 0.5%. In the experiment, since there is little missing data at the station, we choose the temporal nearest neighbor interpolation method to fix the missing value. The interpolation is implemented through python package 'scipy'.

We then need to unify two different data set because they have different time scales. Our experiment requires ozone prediction models for next hour and next day, so we obtain two different temporal resolution data set, namely hourly and daily data sets. We can unify them through the following steps.

*Step*1: For the daily meteorological data in E-OBS, we repeat the daily data 24 times and map it to the hourly air pollution data for that day.

*Step*2: For hourly ozone concentration data, we calculate Mean values of the daily maximum 8-h average (MDA8), which is a commonly used ozone concentration evaluation indicator and has a clear threshold.

*Step*3: For other hourly air pollution data except ozone, we calculate the daily 24-hour average and map it to the daily meteorological data.

Through the above three steps, two data sets on different time scales are obtained. Comparing the size of the two data sets, the hourly data set has 43824 sets of data, while the daily data set has only 1826 sets of data.

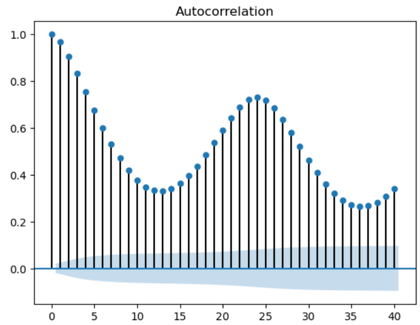
2.3 Data Analysis

We use the autocorrelation function to measure the correlation among ozone on different time scales.

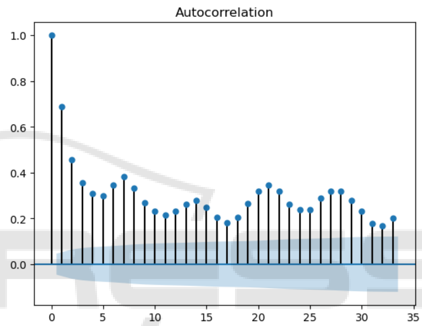
$$\rho_k = \frac{Cov(X(t), X(t+k))}{\sigma_{X(t)}\sigma_{X(t+k)}} \tag{1}$$

where $X(t)$ and $X(t+k)$ represent the ozone of one time series with k time steps difference.

Figures 2 shows the autocorrelation of ozone concentration on different time scales. It can be seen that hourly ozone and daily ozone concentrations are both periodic. The autocorrelation of hourly ozone concentration reaches a peak every 24 time steps, which matches the daily periodicity. Similarly, we can also find that the daily ozone concentration has a weekly



(a) Hourly ozone



(b) Daily ozone

Figure 2: Autocorrelation coefficient of ozone.

cycle. Although the periods are different, the hourly ozone and daily ozone concentrations are both periodic and have the same trend. It provides us with a basis for transfer learning on different time scales.

2.4 Data Transformation in Time Series

In a time series forecast problem, the data is transformed into the following structure.

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_t & x_{t+d} \\ x_2 & x_3 & \cdots & x_{t+1} & x_{t+d+1} \\ x_3 & x_4 & \cdots & x_{t+2} & x_{t+d+2} \\ \vdots & \vdots & \vdots & \vdots & \\ x_n & x_{n+1} & \cdots & x_{t+n-1} & x_{t+n+d-1} \end{bmatrix} \tag{2}$$

where $\{x_1, x_2, x_3, \cdots\}$ is time series data in our experiments, t is the time-step we use in the input and d is the time-step ahead we want to predict. Each row is a sample to be trained in the forecast model, and n rows in the matrix represent a total of n sets of data.

The first t columns are used as input features, and the last column is the output features that need to be predicted. In the experiment, the number of input time steps t is determined by actual problems, while the time step d that needs to be predicted is determined by requirements.

3 METHODS AND EVALUATION

In this section, we introduce the basic principles of the CNN-LSTM and Transfer Learning. At the end of this section, we define the evaluation criteria of the model.

3.1 CNN-LSTM

Convolutional neural network (CNN) is a feedforward network, which is widely used in the fields of natural language processing and image processing. The special structures of CNN are convolutional layers and pooling layers. With these layers, main features of input data are extracted and the parameters required for model fitting will be greatly reduced. It can solve the under-fitting problem when the amount of data is insufficient.

Long short-term memory (LSTM) is a special RNN model. Compared with RNN, LSTM can solve the problem of gradient vanish during long-term sequence training. A common architecture of LSTM is composed of a cell and three gates: an input gate, an output gate and a forget gate. The gates determine whether the information is discarded or retained.

The basic structure of the CNN-LSTM model includes CNN layers, pooling layers, LSTM layers and a final fully connected layer. First, the features of input time series are extracted through the CNN layers, and then the pooling layer is used to retain the main features while reducing the parameters. The LSTM layers are trained to find the dependence between different time steps in the time series. Finally the output is predicted through the fully connected layer.

3.2 Transfer Learning

Transfer learning (TL) is a research problem in machine learning which aims to get knowledge from a different but related problem. It can solve the problem of insufficient data in some fields. The structure of transfer learning can be seen in Figure 3.

Given a learning task based on target domain, we can get knowledge from similar task based on source domain (Tan et al., 2018). Transfer learning can improve the performance of model in target domain, especially when the amount of data is insufficient. It is widely used in image recognition and natural language processing problems. At present, there are few studies on transfer learning related to air pollution prediction.

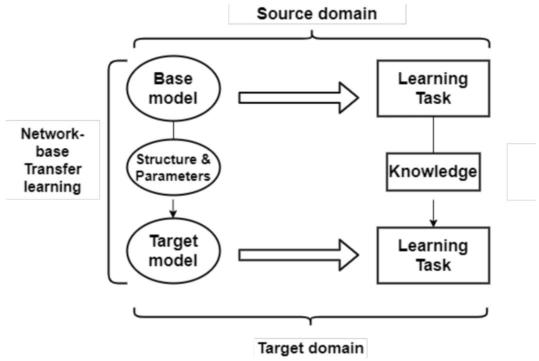


Figure 3: Transfer Learning.

In our experiments, network based deep transfer learning is used (Tan et al., 2018). It can reuse the partial network of basic model that has been trained in the source domain, such as structures and parameters. These information can be transferred into the model in target domain. In network-based deep transfer learning models, front-layers can be treat as feature extractor and the remaining layers are the predictor of target task.

In our experiment, hourly ozone prediction model is used as basic model trained in source domain and daily ozone prediction model is in target domain. Knowledge from small time scale is transferred to a larger time scale. In our CNN-LSTM model, the front CNN layers and part of LSTM layers are retained as feature extractors and additional LSTM layers are added to train the new model.

3.3 Evaluation Criterion

We use root mean squared error (RMSE) and Pearson correlation coefficient to evaluate both the hourly and daily models. For daily ozone prediction model, we also compare the training time of different model, which is also an advantage of transfer learning.

For Mean values of the daily maximum 8-h average (MDA8), there is an evaluation threshold from World Health Organization (WHO), which is 100 $\mu g/m^3$. Ozone exceedances can be quantified through this threshold. There is no formal threshold for low ozone concentration, but previous studies have also found that low ozone concentrations are also harmful to health. Here we select 60 $\mu g/m^3$ as another threshold. In order to verify whether the model can accurately predict ozone to the range it belongs to, we used the above two thresholds to divide the ozone into three categories. Confusion matrix is used to show the result. The approximate ratio of the three categories from low to high is 4:5:1. It means that samples over 100 $\mu g/m^3$ only account for 10% of the total data, which brings difficulties to accurate predictions.

4 EXPERIMENT AND ANALYSIS

In this section, we will introduce in detail the experiments. First, we use CNN-LSTM to predict ozone concentration for next hour and compare the result with other commonly used ozone forecast machine learning models. After that, we use the daily data set to predict MDA8 for next day. In this case, the amount of data in this data set is insufficient, which makes it difficult for the CNN-LSTM model to accurately predict the ozone concentration. Therefore, we investigate using the hourly model as a basic model for transfer learning to improve the forecast result.

4.1 One-hour Ozone Prediction

CNN-LSTM is used to train the model for one-hour ozone forecast. For evaluation, we compare the result of CNN-LSTM to other commonly used ozone prediction machine learning models in previous studies. Only hourly data set is used to train the model.

There are still some hyper parameters about neural networks that need to be determined before model fitting. As described in Section 2.3, the hourly ozone concentration is periodic. The ozone concentration every 24 hours shows a strong autocorrelation. Therefore, we select the data of past 24 hours as input and the ozone concentration for the next hour as output. For each set of inputs, we use all 13 elements as features, including 8 air pollution elements and 5 meteorological elements. The number of layers and neurons in forecast model is determined through cross-validation. We set the number of layers of CNN and LSTM to 1 to 5 and neurons in each layer to 40, 60, 80 and 100, respectively. Through 5-fold cross-validation, 2 CNN layers with 100 neurons per layer and 4 LSTM layers with 60 neurons per layer are finally selected. Dropout layers are used between LSTM layers to prevent overfitting. One max-pooling layer is added to the model after CNN layers. At the same time, we add a fully connected layer containing a neuron before the output layer. Meanwhile, *relu* function is selected as activation function and *adam* is selected as optimizer.

In the hourly data set, we select the first 20,000 sets of data as the training set, and the remaining data as the test set to evaluate the model. We can see the prediction results of the CNN-LSTM model in Figure 4. In Figure 4, we only show the result of last 120 hours in test set. The dotted line is the observation of ozone concentration, and the red line is the prediction result of CNN-LSTM. From the figure, we can see that the red line and the dotted line basically coincide, which means that the CNN-LSTM model can fit the overall trend of ozone concentration well. At the same time, the CNN-LSTM model can also simulate the daily periodic changes of ozone.

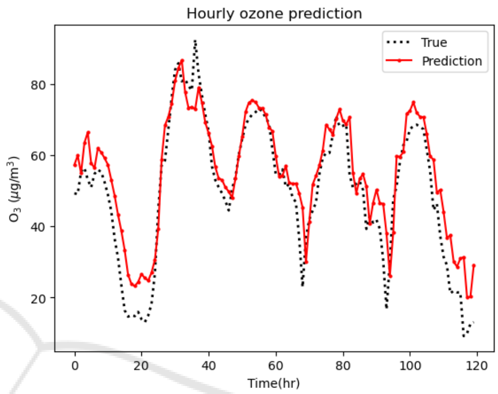


Figure 4: CNN-LSTM for 1-h ozone forecast.

We mainly compared CNN-LSTM with three different models which are commonly used in previous studies, namely, Random Forest (Feng et al., 2019), MLP (Sayeed et al., 2020) and LSTM (Eslami et al., 2019). The RMSE and Pearson correlation coefficient of each model are shown in Table 2.

Table 2: Performance of different model for ozone forecast.

Model	CNN-LSTM	LSTM	MLP	RF
RMSE	9.05	11.91	13.81	13.25
Corr	0.96	0.95	0.95	0.90

It can be seen from Table 2 that all four models can simulate the trend of ozone well. Among them, the prediction result of the random forest method (500 regression trees) has the lowest correlation coefficient with the observation, only 0.90. Comparing the RMSE of four models, we can find that CNN-LSTM can get the smallest RMSE, only 9.05. It can get more accurate prediction results than the other three models.

Besides, we also compared with other ensemble methods, such as gradient boosting (500 regression trees). Their performance is similar to the random forest, so we did not list them in Table 2. We also compared our model with the python package ’Prophet’, which is commonly used in time series prediction problem (Taylor and Letham, 2018). However, since the input for 'Prophet' only contains ozone data, the model can only simulate the general trend of ozone and cannot accurately predict the ozone concentration. Its RMSE is above 20, much larger than our model.

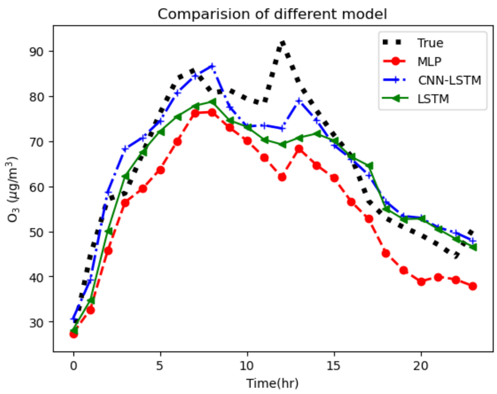


Figure 5: Prediction of different models.

In order to compare the three models with similar results, namely MLP, LSTM and CNN-LSTM, we show the ozone peak part of the day in Figure 5. Although all three models tend to underestimate the daily maximum value of ozone concentration, the CNN-LSTM model still performs the best among them. In comparison, CNN-LSTM has obvious advantages in the prediction of daily ozone peaks.

4.2 Transfer Learning based Daily Ozone Prediction

It is shown in the previous section that CNN-LSTM is effective for 1-hour ozone prediction. In the following, we use daily data set to fit new models for daily ozone prediction. However, a larger time scale will result in greater changes in ozone at adjacent time steps. Meanwhile, less data are available for training. From section 2.2, we can know that the hourly data set has more than 40,000 sets of data, while the daily data set has only 1826 sets of data. In this section, we use houly CNN-LSTM ozone prediction model as a basic model to train the daily ozone prediction model through transfer learning, namely TL-CNN-LSTM ozone prediction model.

For the daily ozone concentration data set, although MDA8 do not have the same cycle as the hourly ozone concentration, the overall trend are similar. At the same time, the CNN-LSTM model can effectively extract the features in the time series. It makes it possible for us to retain features extracted from front-layers of hourly model and use transfer learning to improve the performance of daily model. We select the first 1200 sets of data in the daily data set as the training set, and the remaining data as the test set. In order to show the effect of transfer learning, we keep the parameters the same as the hourly model in the experiment.

In the training process of TL-CNN-LSTM model, one new parameter is the number of frozen layers extracted from the basic model. It determines how many layers of the base model needs to be retained. Starting from the first CNN layer, the parameters in the hourly CNN-LSTM model are retained layer by layer. At the same time, up to 4 LSTM layers are added to the new model after frozen layers. Table 3 shows that when 2 CNN layers and 2 LSTM layers are retained, the prediction result obtained by transfer learning model is the best. They have the least RMSE and the largest correlation coefficient. It suggests that too many or too few retained features can both make the target model perform worse in transfer learning. The result of TL-CNN-LSTM model is shown in Figure 6.

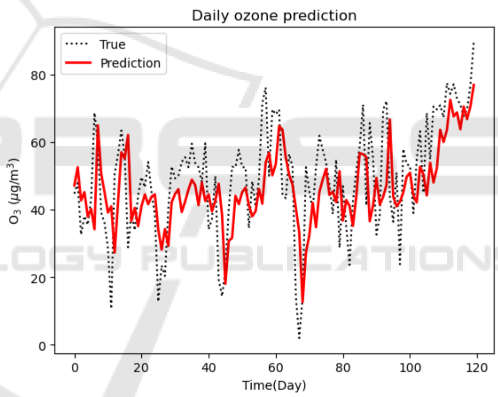


Figure 6: TL-CNN-LSTM for 1-day ozone forecast.

We also compare TL-CNN-LSTM with LSTM and CNN-LSTM, which perform well in the previous hourly forecast models. From Table 4, we can find the negative impact of insufficient data on the above model. All three models have big RMSE in the daily model, while TL-CNN-LSTM performs the best. Compared with CNN-LSTM, the RMSE in the prediction results of TL-CNN-LSTM is reduced by 21%, and the correlation coefficient is increased by 13%. At the same time, the training time of daily model is also greatly reduced through transfer learning. The run time to train TL-CNN-LSTM is only about half of CNN-LSTM. It implies that the features in the hourly CNN-LSTM model can be extracted and transferred by transfer learning, which can improve the predictive performance of the target model with larger time scale.

Table 3: Result of TL-CNN-LSTM with different frozen layers.

Frozen layer	CNN-1	CNN-2	LSTM-1	LSTM-2	LSTM-3	LSTM-4
RMSE	20.54	18.40	15.55	14.92	15.08	16.82
Corr	0.71	0.72	0.80	0.81	0.80	0.74

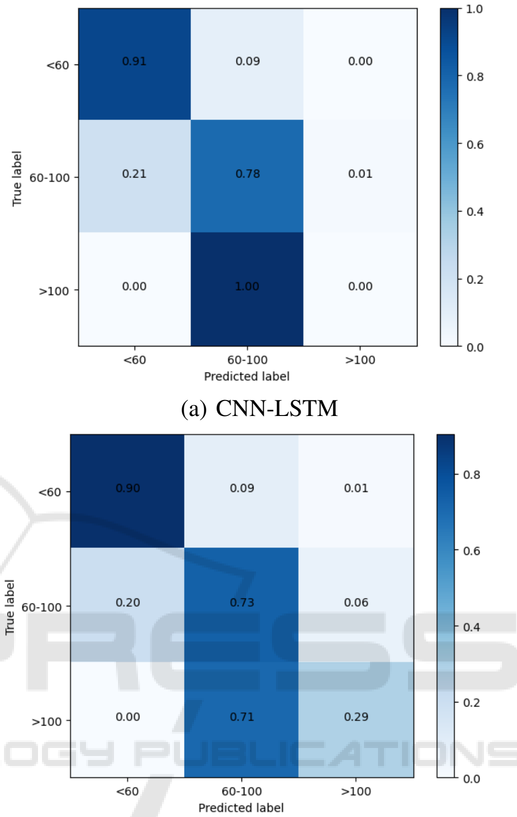
Table 4: Comparison of different model.

Model	LSTM	CNN-LSTM	TL-Model
RMSE	19.14	18.83	14.92
Corr	0.70	0.71	0.81
Time(s)	306	168	75

However, the model still has shortcomings in the daily peak forecasting due to the serious lack of data in this part. As described in Section 3.3, we divide MDA8 in the test set into three classes. We evaluate the model's prediction accuracy for the range of ozone concentration through confusion matrices, which can be seen in Figure 7.

In general, events of high ozone concentration are particularly important which require our attention, because these have direct impact on human health. It can be seen from Figure 7 that for low and medium ozone concentrations, the CNN-LSTM model can already predict its range accurately. Whereas, the CNN-LSTM model do not predict ozone concentrations higher than $100\mu g/m^3$ at all, while TL-CNN-LSTM can accurately predict 29% of this part of the data. As mentioned in section 3.3, only around 10% of ozone data is higher than $100\mu g/m^3$, which means only around 2000 samples in the hourly data set and 100 samples in the daily data set in this class. Neither the transfer learning method nor the CNN-LSTM model can effectively extract relevant features.

For transfer learning, it is difficult to improve the prediction accuracy of extreme events, that is, ozone exceedances in our experiments. For this problem, one possible solution is to simulate more high daily ozone concentration data through methods such as resampling, which has been proven to be effective in visibility prediction (Deng et al., 2019). However, ozone data simulation will be more difficult because the ozone concentration is periodic.



(a) CNN-LSTM

(b) TL-CNN-LSTM

Figure 7: Confusion Matrix.

5 CONCLUSIONS

We use two data sets with different time scale to predict ozone concentration with CNN-LSTM model and network-based transfer learning methods. CNN-LSTM is used to predict ozone concentration for next hour with hourly data set. Compared with three commonly used models in previous studies, namely, RF, MLP and LSTM, CNN-LSTM performs the best in both overall trend and accuracy. In particular, for the prediction of peak ozone concentration, CNN-LSTM performs significantly better than other models. It shows that CNN-LSTM model can be used in hourly ozone prediction problems and performs well.

We also used daily ozone data set to predict ozone for next day. CNN-LSTM model is used and the result is improved through the transfer learning method based on hourly CNN-LSTM model. In a daily time scale, the number of samples is greatly diminished. Insufficient training data leads to the CNN-LSTM model with a large error in daily ozone prediction problem. The network-based transfer learning method can obtain the extracted features from the CNN-LSTM model of a smaller time scale and improves the performance of target model. Compared to daily CNN-LSTM model, TL-CNN-LSTM model can reduce the RMSE by 21% and training time by 55%.

In current practice, we still need to use the traditional chemical transport model to predict ozone, because it is more accurate in case of high ozone concentrations. Compared to the chemical transport model, our TL-CNN-LSTM model is more flexible and can be applied to various local problems, such as ozone concentration prediction at a single site. At the same time, the machine learning method greatly saves the time and resource consumption of model training. However, in the case of ozone exceedances, the severe lack of relevant samples makes even transfer learning model can not predict accurately. Network-based transfer learning only enables the target model to obtain main features from similar models, and a certain amount of data corresponding to cases of interest is still needed to train the new model with different parameters. To solve this problem, we can add more input samples by re-sampling or other methods. In future research, we will investigate adding other ozone-related elements to the input data, such as predicted future temperature, to increase the accuracy of model predictions. At the same time, our current experiment is only based on the data from one site. We will use data from more sites to train and optimize the model through spatial transfer learning.

ACKNOWLEDGEMENTS

We thank the German environmental agency for the air pollution data which is used in the case study for training the neural network models.

REFERENCES

Council, N. R. et al. (1992). *Rethinking the ozone problem in urban and regional air pollution*. National Academies Press.

Curier, R., Timmermans, R., Calabretta-Jongen, S., Eskes, H., Segers, A., Swart, D., and Schaap, M. (2012). Improving ozone forecasts over europe by synergistic use of the lotos-euros chemical transport model and in-situ measurements. *Atmospheric environment*, 60:217–226.

Deng, T., Cheng, A., Han, W., and Lin, H.-X. (2019). Visibility forecast for airport operations by lstm neural network. In *ICAART (2)*, pages 466–473.

Eslami, E., Choi, Y., Lops, Y., and Sayeed, A. (2019). A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing and Applications*, pages 1–15.

Feng, R., Zheng, H.-j., Zhang, A.-r., Huang, C., Gao, H., and Ma, Y.-c. (2019). Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in hangzhou, china. *Environmental pollution*, 252:366–378.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Iglesias, D. J., Calatayud, Á., Barreno, E., Primo-Millo, E., and Talon, M. (2006). Responses of citrus plants to ozone: leaf biochemistry, antioxidant mechanisms and lipid peroxidation. *Plant Physiology and Biochemistry*, 44(2-3):125–131.

Ma, J., Cheng, J. C., Lin, C., Tan, Y., and Zhang, J. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, 214:116885.

McKee, D. (1993). *Tropospheric ozone: human health and agricultural impacts*. CRC Press.

Oh, S. L., Ng, E. Y., San Tan, R., and Acharya, U. R. (2018). Automated diagnosis of arrhythmia using combination of cnn and lstm techniques with variable length heart beats. *Computers in biology and medicine*, 102:278–287.

Placet, M., Mann, C., Gilbert, R., and Niefer, M. (2000). Emissions of ozone precursors from stationary sources:: a critical review. *Atmospheric Environment*, 34(12-14):2183–2204.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., and Jung, J. (2020). Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, 121:396–408.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. In *International conference on artificial neural networks*, pages 270–279. Springer.

Tang, X., Zhu, J., Wang, Z., and Gbaguidi, A. (2011). Improvement of ozone forecast over beijing based on ensemble kalman filter with simultaneous adjustment of initial conditions and emissions. *Atmospheric Chemistry and Physics*, 11(24):12901.

Taylor, S. J. and Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1):37–45.

Wang, J., Yu, L.-C., Lai, K. R., and Zhang, X. (2016). Dimensional sentiment analysis using a regional cnn-lstm model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 225–230.

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M. L., Zhang, M., and Di, B. (2018). Spatiotemporal prediction of daily ambient ozone levels across china using random forest for human exposure assessment. *Environmental Pollution*, 233:464–473.

Zhao, R., Yan, R., Wang, J., and Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, 17(2):273.