Data assimilation model based on machine learning

Jialin Lang¹*, Feng Qiu¹ and Pin Wu¹
¹School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China
*Corresponding author’s e-mail: alang_child@shu.edu.cn

Abstract. Data assimilation (DA) is a method mainly absorbs the observation data into the simulation model, integrates the errors of observation and simulation, and provides a more accurate state so as to reduce the forecast error. Data assimilation has been widely used in the fields of atmosphere and ocean. However, traditional data assimilation methods require a lot of computing resources and consume a long time. Machine learning is a data analysis method with strong learning ability and rapid prediction ability. Long Short-Term Memory network (LSTM) is a widely used machine learning model, which has good effect on time series prediction. In this paper, we use the historical data of data assimilation to train LSTM model and get the prediction model. The experimental results show that the LSTM model can learn the latent law from the historical data, the results of the model fit the real data well, and the calculation speed is greatly improved compared with the original data assimilation algorithm.

1. Introduction

1.1. Data assimilation

Data assimilation (DA) integrates new observation data during the operation of numerical models on the basis of considering the spatial and temporal distribution of data and the errors of observation and background. The traditional data assimilation algorithms are complicated, and have the problem of large amount of computation and long computation time. However, the running time of data assimilation determines whether this process can be carried out effectively. If the calculation time is too long, the calculation result will be no longer of practical significance. Figure 1 is a traditional data assimilation and prediction model.

Figure 1. Data assimilation process. Traditional DA algorithm is complicated and time consuming.

1.2. 4Dimensional Variational Algorithm

3dimensional variational algorithm (3D-Var)¹ and 4dimensional variational algorithm (4D-Var)² are two classical data assimilation methods. The basic idea of the variational assimilation method can be summarized as follows: the observation data are optimally assimilated into the background field to obtain new system state variables. To do this, we need to establish the objective function and minimize it. In the 3D-Var, the objective function is as follows:
\[ J(X) = (X - X^b)^T B^{-1} (X - X^b) + (Y - HX)^T R^{-1} (Y - HX) \]  

(1)

In the function, \( J(X) \) is the cost function, \( X \) is the state variable, \( X^b \) is the background state variable, \( B \) is the background error covariance matrix, \( Y \) is the observation data, \( H \) is the observation operator, \( R^{-1} \) is the observation error covariance matrix. The state variable that minimizes the value of \( J(X) \) is the optimal estimated state variable.

The time interval of 3D-Var analysis is equal to the observation period, but in fact, the observation period of observation sources may be different, 4D-Var takes different time intervals into account to avoid losing useful information and processes these observations at each observation moment. The objective function is as follows:

\[
J(X) = (X - X^b)^T B^{-1} (X - X^b) + \sum_{k=d}^T (Y_k - H_k (M_k (\cdots (M_1 (X)))))^T R^{-1} (Y_k - H_k (M_k (\cdots (M_1 (X)))))) 
\]

(2)

\( M \) is the change of state variable with time, the other parameter definition is the same as 3D-Var.

It's obvious that the computational complexity of this method is greatly increased. But efficiency is critical to the data assimilation process.

2. Our model

2.1. Long Short-Term Memory network

Artificial Neural Network (ANN) has been a hot topic in the field of artificial intelligence since the 1980s. It abstracts the neuron of human brain from the perspective of information processing, and forms different networks according to different connection modes. ANN has made great progress and solved many practical problems in the fields of intelligent robot, automatic control, medicine and so on[3]. In recent years, people have tried to apply machine learning methods to data assimilation[4][5].

Recurrent Neural Network (RNN)[6] is a kind of Artificial Neural Network with short-term memory. In the network, the neuron can not only accept the information of other neurons, but also its own information to form a network structure with loops. In practice, RNN will have the problem of gradient vanishing or gradient explosion, which makes it difficult to deal with Long-Term Dependencies problem. To solve this problem, gating mechanism is introduced, and long short-term memory network(LSTM) is one of the solutions. Because of its unique design, LSTM has been widely used in natural language processing (NLP) tasks and Speech Recognition, and has achieved good results[7][8][9].

2.2. Model design

Since the data assimilation algorithm is complex and rich in historical data, and LSTM can learn hidden laws from long-term historical data, then we can train the LSTM model with the historical data of data assimilation. The trained model should be able to accurately perform, more than that, the calculation efficiency is far better than traditional data assimilation algorithms. Figure 2 is the schematic of the model we want to get.

![Figure 2. Our model. We use historical data to train LSTM model to replace traditional process.](image-url)
3. Experiment

3.1. Dataset
Lorenz63 system is a simple mathematical model of atmospheric convection established by Lorenz in 1963\(^{[10]}\). It is the earliest numerical representation of a chaotic system. Chaotic system refers to the existence of seemingly random irregular motion in deterministic system, which is very sensitive to the change of initial value. Therefore, Lorenz63 system is often used to study data assimilation methods. The governing equation of the system is as follows:
\[
\begin{aligned}
\frac{dx}{dt} &= -\sigma x + \sigma y \\
\frac{dy}{dt} &= \gamma x - y - xz \\
\frac{dz}{dt} &= xy - bz
\end{aligned}
\] (3)

This governing equation can describe the convective motion of the fluid. Where \(\sigma\) and \(\gamma\) denote Prandtl number and Rayleigh number respectively, \(b\) is the parameter associated with convection scale. \(x, y, z\) denote fluid velocity, horizontal temperature difference and vertical temperature difference respectively. The equation can be solved by Runge-Kutta numerical integration method. Figure 3 is the operating state of the Lorenz63 system under a certain initial condition.

![Figure 3. The simulation data of Lorenz63. One point represents the state of the system at a moment.](image)

For the purpose of experiment, the Lorenz63 system is used as the chaotic system to calculate the operating state of the system under a certain initial condition, and the random noise is added to simulate observation values. The random noise follows the Gaussian distribution. And then, the 4D-Var is used for data assimilation, in the end, the obtained results are taken as the dataset.

3.2. Experiment setup
The experiment used an Intel(R) Core(TM) i5-8300h processor and an NVIDIA GeForce GTX1050 graphics card, and the network model was built with keras2.3.1, which is convenient and stable. We used the mean square error(MSE) as the loss function, and used the stochastic gradient descent(SGD) method for parameter training. Ten thousand pieces of data were used in each experiment, of which 9 thousand pieces were used for training and 1 thousand pieces were used for verification. The number of trainable parameters of the model was about \(10^4\).

3.3. Results and analysis
In order to ensure the reliability of the experiment, we set different initial states of the chaotic system and carried out many experiments. Table 1 shows the results of several experiments.
Table 1. The results of several experiments.

| Experiment | Time steps of forecast | Min error | Max error | Average error | Time consumption for LSTM(S) | Time consumption for 4D-Var(S) |
|------------|------------------------|-----------|-----------|---------------|-------------------------------|-------------------------------|
| 1          | 1000                   | 1.15e-12  | 0.0225    | 1.93e-4       | 0.386                         | 52.376                        |
| 2          |                        | 4.75e-10  | 0.0188    | 1.87e-4       | 0.379                         | 53.477                        |
| 3          |                        | 5.21e-9   | 0.0287    | 1.93e-4       | 0.369                         | 53.034                        |
| 4          |                        | 7.46e-10  | 0.0226    | 1.70e-4       | 0.405                         | 53.188                        |
| 5          |                        | 8.94e-10  | 0.0144    | 1.95e-4       | 0.388                         | 52.957                        |
| Average    |                        | 1.465e-9  | 0.0214    | 1.85e-4       | 0.390                         | 52.985                        |

The figure 4 is the prediction results of an experiment. It can be seen from the figure that the experimental results fit the original values well, indicating that the method has a certain reliability.

4. Conclusion

In this paper, in order to make full use of the historical data of data assimilation and improve the efficiency of data assimilation, we construct the LSTM model and train it with the historical data of data assimilation. The results show that the LSTM model can effectively learn the potential laws in the data, and the model obtained by the training can accurately predict the data assimilation results in a certain period of time, and its computational efficiency is much faster than that of 4D-Var. In the future, we will try to apply the method to more practical applications to test it.

References

[1] Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., & Hollingsworth, A., et al. (1998). The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: formulation. Quarterly Journal of the Royal Meteorological Society.

[2] Courtier, P., & Talagrand, O.. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. ii: numerical results. Quarterly Journal of the Royal Meteorological Society.

[3] Hinton, G., Vinyals, O., & Dean, J.. (2015). Distilling the knowledge in a neural network. Computer ence, 14(7), 38-39.

[4] B, P. W. A., B, X. C., B, J. S., B, W. Z., C, R. A., & B, Y. G. C.. (2021). Fast data assimilation (fda): data assimilation by machine learning for faster optimize model state. Journal of Computational Science.

[5] Arcucci, R., Zhu, J., Hu, S., & Guo, Y. K.. (2021). Deep data assimilation: integrating deep learning with data assimilation. Applied Sciences, 11(3), 1114.

[6] Mikolov, T., Martin Karafiát, Burget, L., Cernock, J., & Khudanpur, S.. (2015). Recurrent neural
network based language model. Interspeech, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September. DBLP.

[7] Yu, D., & Deng, L.. (2014). Automatic Speech Recognition: A Deep Learning Approach. Automatic speech recognition : a deep learning approach. Springer.

[8] Zhu, C., Chen, F., Ahmed, U., & Savvides, M.. (2021). Semantic relation reasoning for shot-stable few-shot object detection.

[9] Du, Y., Yan, Y., Chen, S., & Hua, Y.. (2020). Object-adaptive lstm network for real-time visual tracking with adversarial data augmentation.

[10] Lorenz, & Edward, N.. (2004). Deterministic nonperiodic flow. Journal of Atmospheric Sciences, 20(2), 130-141.