Hybrid data assimilation based on multilayer perceptron

Jialin Lang* and Feng Qiu

1School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

*Corresponding author’s e-mail: alang_child@shu.edu.cn

Abstract. Data assimilation is widely used in weather forecasting, ocean forecasting, remote sensing observation and other fields. The result of data assimilation directly affects the accuracy of forecasting. Data assimilation algorithms have been studied extensively. The classical algorithms include 3D-Var, 4D-Var, Kalman filter and ensemble Kalman filter. New assimilation algorithms also emerge one after another. In recent years, people have tried to use a hybrid of various data assimilation methods. Deep learning is a hot research field in recent years. Deep learning is able to mine the hidden features of data for analysis and prediction. In this paper, we attempt to use the classical deep learning model-multilayer perceptron to extract the characteristics of traditional data assimilation algorithms to perform hybrid data assimilation. The experiments show that the results of this method are significantly improved compared with the traditional methods, and reveal the value of multilayer perceptron in the direction of hybrid data assimilation.

1. Introduction

1.1. Chaotic system

A chaotic system refers to a deterministic system where there are seemingly random irregular motions, and its behavior is uncertain, unrepeatable, and unpredictable, that is, chaotic phenomena. Chaotic system is very sensitive to the change of initial value. A small change at the beginning can make a big difference in the future. Because of this, weather forecasting and ocean forecasting are extremely difficult. In order to integrate real-time observation information into the current forecast information, people have carried out data assimilation methods. Among them, the classical and widely used data assimilation methods include variational data assimilation and Kalman filter. Figure 1 illustrates the importance of data assimilation.

![Figure 1. The importance of data assimilation. The prediction curve quickly deviates from the real data without assimilation. The prediction curve remains close to the real data with assimilation.](image-url)
1.2. variational data assimilation
The variational algorithm constructs a cost function to describe the difference between the analysis value and the true value of the state, and uses the variational idea to transform the data assimilation problem into an extreme value solving problem.

The three-dimensional variational algorithm (3D-Var) uses all the observation data in the time window to adjust the model trajectory, and finally fits the model simulation trajectory to all the observation values in the time window.

The four-dimensional variational algorithm (4D-Var) was proposed by Talagrand in 1987. Four-dimensional refers to the three-dimensional distribution of state in space and the one-dimensional time distribution. It is developed on the basis of 3D-Var and makes up for the deficiency of 3D-Var in the time variation of state to a certain extent. The cost function of 4D-Var is as follows:

\[ J(X) = (X - X^b)^T B^{-1}(X - X^b) + \sum_{k=0}^{T-1} (Y_k - H_k(M_k(\cdot\cdot(M_k(X))))^T R^{-1}(Y_k - H_k(M_k(\cdot\cdot(M_k(X))))) \tag{1} \]

In the function, \( M \) represents the change relationship of the state with time, which can better reflect the complex nonlinear constraint relationship.

1.3. Kalman Filter
The Kalman filter algorithm realizes data assimilation in two steps: prediction and update. The prediction step predicts the state of time \( k+1 \) based on the state at the current time \( k \), and the update step adjusts the predicted state value at time \( k+1 \) under the condition of the existing observation data, so as to obtain the optimal estimate of the state at time \( k+1 \). Then, the model is reinitialized with the estimated value of the state at time \( k+1 \), and the above steps are repeated until all the observation data have been used to complete the prediction and update.

The ensemble Kalman filter algorithm (EnKF) \([4]\) calculates the state prediction error covariance by Monte Carlo method \([5]\), and solves the difficulty of estimating and forecasting the background error covariance matrix in the practical application with the idea of ensemble, which can be used for the data assimilation of the nonlinear system, and reduces the computation cost effectively.

2. Related work
Variational data assimilation and Kalman filter have their shortcomings, so some researchers thought of combining the two to maximize their strengths and avoid weaknesses. The idea of hybrid data assimilation \([6]\) was proposed and further developed in the field of data assimilation. In the implementation of hybrid data assimilation, some people directly use fixed coefficients to combine. Obviously, this method cannot meet the various changes of chaotic systems in practical applications. Therefore, we need a more flexible method for hybrid.

3. Our model
Deep learning \([7]\) is a hot research direction in recent years. It learns the hidden laws and information of sample data. The information obtained in the learning process is of great help to the interpretation of data such as text, images and sounds. Deep learning has achieved rich results in machine translation \([8]\), speech recognition \([9]\), image recognition \([10][11]\) and other fields.

Multilayer perceptron (MLP) \([12]\) is a simple artificial neural network. It is characterized by multiple layers. The first layer is called the input layer, the last layer is called the output layer, and the middle layer is called the hidden layer. There is no fixed number of hidden layers in the multilayer perceptron, and there is no limit to the number of neurons in each layer, so you can choose according to your needs.

This paper uses rich historical data of traditional data assimilation to train an MLP model. The trained model should be able to calculate the result after hybrid based on the input of 4D-Var and EnKF. The model we expect to get is shown in the figure 2.
4. Experiment

4.1. Dataset
The chaotic system in this experiment is the lorenz63 system. Figure 3 shows the state of the Lorenz63 system. We calculate its operating state, and use the state as the real data, then add random noise to simulate the observations. The random noise conforms to the Gaussian distribution.

4.2. Experiment setup
The experiment was carried out with an Intel(R) Core (TM) i5-8300h processor and an NVIDIA GeForce GTX1050 graphics card, and the MLP model was built with keras2.3.1, which is stable and convenient. The mean square error (MSE) was used as the loss function, and the stochastic gradient descent (SGD) method was used for parameter training. 90 thousand pieces of data were used in each experiment, of which 80 thousand pieces were used for training and 10 thousand pieces were used for verification. The number of trainable parameters of the model was about 1.2*10⁴.

4.3. Results and analysis
After trial and adjustment, we find out that the ‘tanh’ activation function is the most suitable activation function, and the model size has been adjusted to 7 hidden layers, each with 50 neurons. Table 1 shows the results of several experiments.
Table 1. The results of several experiments.

| Experiment | 1       | 2       | 3       | 4       |
|------------|---------|---------|---------|---------|
| MSE of 4D-VAR       | 3.80e-3 | 1.75e-2 | 1.58e-3 | 1.39e-3 |
| MSE of EnKF         | 2.01e-2 | 7.91e-2 | 1.34e-2 | 1.17e-2 |
| MSE of hybrid model | 6.60e-4 | 1.68e-3 | 1.42e-3 | 1.15e-3 |

The figure 4 below is the result of an experiment. It can be seen from the figure that, compared with the 4D-Var and EnKF, the assimilation result is closer to the true value after using the MLP model for hybrid data assimilation, proving that our model is effective.

![Figure 4](image-url)

Figure 4. A part of the result of an experiment. The black line, red line, green line and blue line represent the real value, hybrid assimilation, 4D-Var, and EnKF result, respectively. In most cases, the result of hybrid assimilation is better than the other two methods.

5. Conclusion

In this paper, a MLP model is used to achieve the hybrid data assimilation of 4D-Var and EnKF algorithm. The results show that the MLP model can learn the underlying laws in the data generated by the two methods. Compared with the two traditional methods, the results of the obtained model are closer to the true value, which improves the accuracy. In addition, new data assimilation algorithms are emerging one after another, and whether more new algorithms can be used for hybrid is a direction worth trying in the future.

References

[1] Parker, T., & Chua, L.. (1992). Practical numerical algorithms for chaotic systems. SPRINGER.
[2] Li, Y., Navon, I. M., Courtier, P., & Gauthier, P.. (1993). Variational data assimilation with a semi-lagrangian semi-implicit global shallow-water equation model and its adjoint. Monthly Weather Review, 121(6).
[3] Welch, G., & Bishop, G.. (1995). An Introduction to the Kalman Filter. University of North Carolina at Chapel Hill.
[4] Evensen, G.. (2009). Ensemble methods. Springer Berlin Heidelberg.
[5] Hastings, W. K.. (1970). Monte carlo sampling methods using markov chains and their applications. Biometrika(1), 97-109.
[6] A., M., Clayton, A., C., & Lorenc. (2013). Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. Quarterly Journal of the Royal Meteorological Society.
[7] Goodfellow, I., Bengio, Y., & Courville, A.. (2016). Deep Learning. The MIT Press.
[8] Ruiter, D., Genabith, J. V., & Espaa-Bonet, C.. (2020). Self-Induced Curriculum Learning in Self-Supervised Neural Machine Translation.
[9] Zhang, L., & Sun, X.. (2021). Study on speech recognition method of artificial intelligence deep learning. Journal of Physics: Conference Series, 1754(1), 012183 (5pp).
[10] Lawrence, S., Giles, C. L., AC Tsoi, & AD Back. (1997). Face recognition: a convolutional neural network approach. IEEE Transactions on Neural Networks, 8(1), 98-113.
[11] Jiang, F., Zhi, X., Ding, X., Tong, W., & Bian, Y. (2020). DLU-Net for Pancreatic Cancer Segmentation. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1024-1028.
[12] M.W, Gardner, and, S.R, & Dorling. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environment.