URSI AP-RASC, Sydney, Australia, 17 – 22 August 2025

Optimized Atmospheric Forecasting Model: Integrating Informer-GRU on IGRA Data

Yuxuan Wang * (1), Jiangting Li (1)
(1) Xidian University, Xi 'an, Shaanxi., China, e-mail: 25201110143@xidian.edu.cn ; jtli@xidian.edu.cn

Abstract

With the advancement of artificial intelligence, atmospheric numerical prediction is increasingly shifting from traditional physical models to deep learning models to address complex data dependencies. This study introduces a novel prediction method that integrates an Informer-GRU model. The model leverages long-distance dependencies in time series through a multi-head attention mechanism and positional encoding to enhance prediction accuracy. Experimental results demonstrate that our model significantly outperforms traditional models in predicting multi-dimensional atmospheric data, achieving approximately 10% reduction in mean squared error. This study not only demonstrates a robust integration of Informer-GRU with Bayesian optimization but also sets a new benchmark for atmospheric numerical prediction, potentially enhancing decision-making in weather-sensitive sectors.

1. Introduction

The Earth's atmosphere is a dynamic system influenced by various meteorological parameters. Understanding and predicting accurately these parameters are crucial for comprehending the behavior and phenomena of the atmosphere. To enhance prediction accuracy, the deep learning method is introduced, offering a new paradigm shift for analyzing atmospheric data. This kind of approach has significantly contributed to the field of atmospheric data prediction and analysis in recent years. In the domain of time series prediction, deep learning models excel by effectively capturing complex temporal dependencies and nonlinear patterns inherent in atmospheric data. This capability significantly enhances the accuracy and reliability of forecasts, particularly in dynamic and multifaceted meteorological conditions. Consequently, these models facilitate a more nuanced understanding of atmospheric phenomena, which can be directly applied to improve operational forecasting systems. The integration of these advanced predictive models allows for a more precise identification and analysis of atmospheric parameters, thereby enriching the practical applications of meteorological research. Deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs) and their variants (e.g., long and short-term memory networks, LSTMs) have become important tools in numerical atmospheric prediction.

2. Dataset

The dataset used in the experiment is the observation data from a single field station of IGRA(Integrated Global Radiosonde Archive). The time span is from January 1, 2000, to December 30, 2020, a total of 7670 days of observation data. Every day, the UTC time data of the IGRA dataset 00:00 is used. The IGRA dataset has 16 dimensions of data, which can be divided into 4 dimensions of data for 4 altitudes. Each height possesses data for temperature, altitude, dew point difference, and pressure variables. In order to verify the robustness and generalization ability of the model on different datasets, the temperature (T), dew point difference (Dpd), and pressure (P) at the lowest level of the IGRA are selected as the dimensions of the model inputs, respectively. The raw data for the IGRA site has long continuous missing values for the period 2004-2007 and shorter missing values for the remaining years (totaling 772 days). In practice, the locations of missing data are first marked, and then these missing values are filled by linear interpolation to obtain complete time series data.

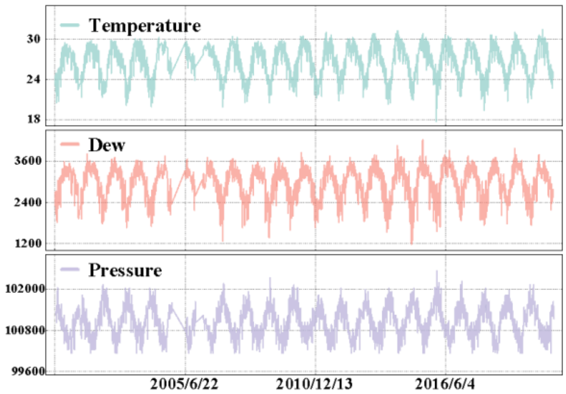

Figure 1. IGRA linear interpolation. The image displays IGRA data that has been linearly interpolated. It features variables such as temperature dew point difference and pressure over a period of 21 years, encompassing a total of 7,760 days.

3. Model

In our model, we introduce GRU layers after both the encoder and decoder blocks. These GRU layers are specifically designed to refine the sequential representation by capturing non-smooth and nonlinear dependencies, which are common in atmospheric data such as temperature and pressure variations. The GRU is particularly suitable for handling sequential data due to its internal gating mechanisms, which help mitigate the vanishing gradient problem during training. The Informer-GRU model developed in this study not only demonstrates the efficacy of integrating GRUs with self-attention mechanisms for multiscale time series analysis but also enhances the model's sensitivity and accuracy in processing temporal information through the strategic use of positional encoding and causal convolution. These innovative features ensure that our model exhibits exceptional performance in predicting atmospheric data, significantly outperforming traditional models.

The model optimizes the computational efficiency and accuracy of time series forecasting mainly through Attention and Causal Convolution. First, the input time series data will be embedded into a high-dimensional vector space, and its time series information will be preserved by Positional Encoding. Then, these embedded data are processed through causal convolutional layers to extract local time-series features, and in the Encoder part, the model extracts global features through a multi-layer multi-head attention mechanism and feed-forward neural network. Each layer of the multi-head attention mechanism enhances the feature extraction capability of the model by computing the attention scores of different subspaces in parallel. Subsequently, in the Decoder section, which has a similar structure to the encoder, the features are further processed through the multilayer multi-head attention mechanism and feed-forward neural network. Eventually, the output of the decoder will be mapped to the target dimension through the fully connected layer to obtain the final prediction result. The specific computational flow is shown in Figure 2.

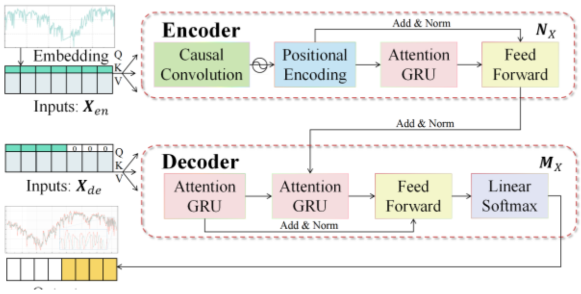

Figure 2. Model Calculation Process. The model adopts an encoder-decoder structure, utilizing positional encoding and attention mechanisms, combined with GRU units to capture the multi-scale nonlinear characteristics of time series.

4. Experiment

This section describes the experimental setup, the specific implementation of the control model, the evaluation metrics, and the analysis of the experimental results. In order to comprehensively evaluate the prediction performance of the Informer model, we selected the LSTM model and GRU as the control models. LSTM is a classical deep learning model for processing time series data, which is good at capturing long-term dependencies. GRU is a simplified version of a recurrent neural network. The comparison experiments use the same training data and test data to ensure comparable results.

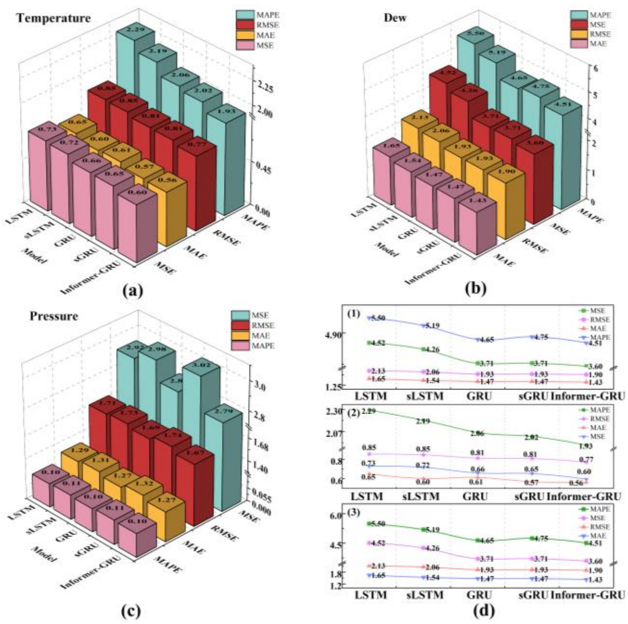

Figure 3. Error statistical result(a) Model Temperature Error Comparison (b) Model Dew Error Comparison (c) Model Pressure Error Comparison (d) Model Error Statistics

Table 1: Error Statistics For Each Model

Variant	model	MSE	RMSE	MAE	MAPE
Tempt	**LSTM**	0.729	0.854	0.653	2.291
	sLSTM	0.718	0.847	0.603	2.187
	GRU	0.659	0.812	0.614	2.061
	sGRU	0.653	0.808	0.570	2.021
	I-G	**0.595**	**0.771**	**0.562**	**1.930**
Dew	**LSTM**	4.156	2.125	1.648	5.496
	sLSTM	4.261	2.064	1.539	5.194
	GRU	3.712	1.927	1.472	4.647
	sGRU	3.713	1.926	1.473	4.750
	I-G	**3.595**	**1.896**	**1.429**	**4.509**
Pressure	**LSTM**	2.915	1.707	1.290	0.101
	sLSTM	2.978	1.726	1.314	0.105
	GRU	2.842	1.686	**1.271**	**0.100**
	sGRU	3.021	1.738	1.324	0.105
	I-G	**2.792**	**1.670**	1.273	0.102

5. Results

In the temperature prediction task, the Informer-GRU model achieved an MSE of 0.595, an RMSE of 0.771, and a MAE of 0.562. These metrics are superior to those of the other models, especially LSTM (MSE of 0.729, RMSE of 0.854, and MAE of 0.653) and GRU (MSE of 0.659, RMSE of 0.812, and MAE of 0.614). Specifically, the Informer-GRU model has an MSE that is 18.4% lower than the LSTM and 9.7% lower than the GRU, an RMSE that is 9.7% lower than the LSTM and 5.1% lower than the GRU, and an MAE that is 13.9% lower than the LSTM and 8.5% lower than the GRU.

In the dew-point difference prediction task, the MSE of the Informer-GRU model is 3.595, the RMSE is 1.896, and the MAE is 1.429. The performance of the Informer-GRU model is also significantly better compared to the LSTM (MSE is 4.516, RMSE is 2.125, and MAE is 1.648) and the GRU (MSE is 3.712, RMSE is 1.927, and MAE is 1.472), the performance of the Informer-GRU model is equally significant. Specifically, Informer-GRU's MSE was 20. 4% lower than LSTM and 3.2% lower than GRU; RMSE was 10. 7% lower than LSTM and 1.6% lower than GRU; and MAE was 13. 3% lower than LSTM and 2.9% lower than GRU. These results show that the Informer-GRU model can effectively reduce the prediction error in predicting the difference between points of dew.

In the pressure prediction task, the Informer-GRU model has an MSE of 2.792, an RMSE of 1.670, and an MAE of 1.273. Although its performance improvement is slightly less pronounced compared to the temperature and dew point difference tasks, it still outperforms both LSTM (with an MSE of 2.9154, an RMSE of 1.707, and an MAE of 1.290) and GRU (with an MSE of 2.842, RMSE of 1.686, and MAE of 1.271). Specifically, the Informer-GRU model has an MSE that is 4.2% lower than LSTM and 1.8% lower than GRU, an RMSE that is 2.2% lower than LSTM and 0.9% lower than GRU, and an MAE that is 1.3% lower than LSTM and slightly higher than GRU but still competitive. The Informer-GRU model performs moderately well with pressure time series data but does not improve as much as it does with temperature and dew point differences. This may be due to the high volatility of the pressure data, which makes it difficult for the model to capture the complex patterns in it. It may also be related to the parameter selection and training process of the model, which needs to be further optimized and adjusted.

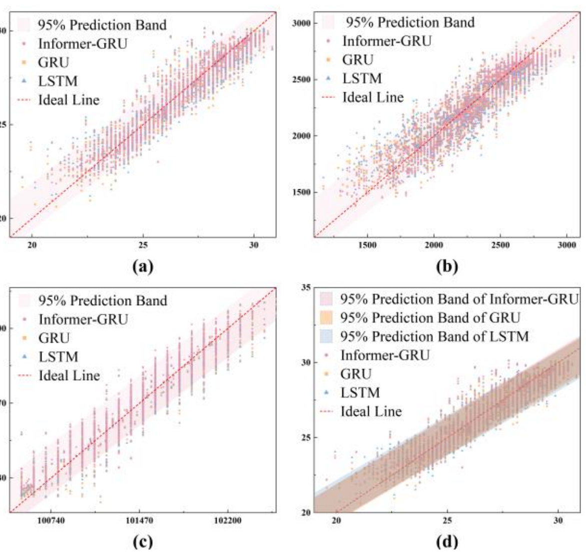

Figure 4. (a) Scatterplot of temperature prediction (b) Scatterplot of dew point difference prediction (c) Scatterplot of pressure prediction (d) Plot of confidence interval comparison

These results show that the Informer-GRU model performs better in capturing complex patterns in temperature time series. The scatter chart of prediction results is shown in Figure 4. The results show that the Informer-GRU model has significant advantages in dealing with nonlinearities and long-term dependencies, resulting in more accurate and stable forecasts.

References

[1] B. Zhao, H. Lu, S. Chen, J. Liu, D. Wu, Convolutional neural net- works for time series classification, Journal of systems engineering and electronics 28 (2017) 162–169.

[2] G. Lai, W.-C. Chang, Y. Yang, H. Liu, Modeling long-and short-term temporal patterns with deep neural networks, in: The 41st international ACM SIGIR conference on research & development in information re- trieval, 2018, pp. 95–104.

[3] S. Ma, T. Zhang, Y.-B. Zhao, Y. Kang, P. Bai, Tcln: A transformer- based conv-lstm network for multivariate time series forecasting, Applied Intelligence 53 (2023) 28401-28417.

[4] R. Wan, S. Mei, J. Wang, M. Liu, F. Yang, Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting, Electronics 8 (2019) 876.

[5] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, In- former: Beyond efficient transformer for long sequence time-series fore- casting, in: Proceedings of the AAAI conference on artificial intelligence, volume 35, 2021, pp. 11106–11115.

[6] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, et al., Time-llm: Time series forecasting by reprogramming large language models, arXiv preprint arXiv:2310.01728 (2023).