Application of a Long Short-Term Memory Recurrent Neural Network to Aviation Weather Forecasting

Chieh-Ni Huang[1], Chuen-Jyh Chen[2]

[1]Department of Aeronautics and Astronautics, National Cheng Kung University, Tainan, Taiwan

[2]Department of Aviation and Maritime Transportation Management, Chang Jung Christian University, Tainan, Taiwan

Abstract

Weather forecast is vital to aviation safety. Inaccurate forecast not only causes problems to pilots and air traffic controllers, but also leads to aviation accidents and incidents. To enhance the forecast accuracy, a deep learning recurrent neural network by long short-term memory is proposed in this work. A dataset of 10 weather features of the whole year of 2020 in Taiwan was employed in the deep learning network, 90% of the data for training and 10% for testing. The data are first standardized in preprocessing and then trained by the long short-term memory network. The results show that the feature of sea pressure, dew point, relative humidity visible mean and cloud amount are most influential for predicting temperature, compared with the features of wind speed, wind direction, sunshine rate and global solar radiation. Deleting the less influential features can achieve the lowest root mean square error (RMSE) to 2.3872, while expediting the forecast calculation. The deep learning network is shown effective to forecast the aviation weather with the mean absolute percentage error (MAPE) 10.8682%.

Key word: deep learning, long short-term memory, recurrent neural network, aviation weather.

1. Introduction

Accurate weather forecast is known important to aviation safety and its chaotic nature has posed many challenges to scientists around the world. Bad weather is the primary feature known to influence on aircraft operation, increase operating cost, and causes accidents. In aviation safety, detail observation of atmosphere states (features), such as temperature, pressure, relative humidity, wind speed and direction, cloud height and density and rainfall have been adopted in weather forecast. For example, at a certain pressure with high temperature, an aircraft needs higher take off speed and requires longer runway.

Numerical weather prediction (NWP) by using mathematic models was first attempted over a century ago [1]. It was not until 1950, NWP was seen in actual applications [2]. Output statistics (MOS) has been most widely used to improve forecast ability [3], but it is being phased out by recent development on artificial neural network (ANN). Schizas et al. [4] applied the backpropagation algorithm in ANN to temperature forecast. Maqsood et al. [5] presented an ensemble of ANN with learning paradigms for weather forecasting. However, ANN still have two serious problems, one is to capture sequential information from the input and the other is the vanishing and exploding gradient problem leading to poor convergence.

Recurrent Neural Network (RNN) by using a looping constraint on the hidden layer of ANN was proposed to solve the difficulties when processing sequential data. Hopfield [6] also developed a single-layer feedback neural network to solve the optimization problems. Based on his work, an RNN with recurrent connection on the hidden state by using backpropagation algorithm was proposed for deep learning [7]. Vlachas et al. showed that RNN trained via Backpropagation through time (BPTT) has superior forecasting abilities and capture well the dynamics of reduced order systems [8]. RNN has the advantage of sharing the parameters across different time steps for effective processing in sequential, or time series, data. There are some weather modelling applications of time series on aviation meteorology or geoscience. Using Bayesian statics method to learn long time series data from observations of high-dimensional chaotic dynamics such as geophysical flows [9]. Illustrate the equivalences between four-dimensional variational data assimilation and an RNN, addressing the grand challenge of making better use of observations to improve physical models of earth system processes [10], and been used successfully for emulating, downscaling, and forecasting weather and climate processes [11]. To gain further process understanding of Earth system science problems,

improving the predictive ability of seasonal forecasting and modelling of long-range spatial connections across multiple timescales [12]. However, when applied to geosciences, an additional difficulty arises by the continually increasing sophistication of the environmental models, so far have not delved into its conceptual and methodological complexities [13].

This work aims to apply a long short-term memory network (LSTM) to model a time series data and to reach high accuracy prediction in weather forecast. For aviation safety, if various weather phenomena can be accurately classified and processed successfully, accurate prediction is achievable for pilot decision making.

2. Long Short-Term Memory Network

An artificial neural network (ANN) is constructed by assuming the input do not have any order and the output depend only on the input. In practice, the output is often dependent upon the previous input and RNN becomes needed. RNN has been known suffering from the problem of vanishing and exploding gradient, where error is accumulated during update (iteration), leading to very large gradients and eventually divergence, i.e., unable to learn from the training data such that a network fails to learn from the input sequential data. A LSTM in RNN is therefore proposed to solve the gradients problems in sequence prediction [14]. Salman et al. [15] proposed a forecasting model extended from LSTM predict weather data in Indonesia airport. Hong et al. [16] adapted an LSTM to very short-term weather forecasting. With the development of deep learning, LSTMs are now well-suited to classifying, processing and predicting the long-term time series data. Sangiorgio et al. [17] show that LSTM architectures maintain good performances when the number of time lags included in the input differs from the actual embedding dimension of the dataset. Vlachas et al. [18] proposed a hybrid architecture to extend the forecasting capabilities of LSTM networks. The handwriting recognition for classifying [19]. The ozone data for processing [20], and the weather data for predicting [21] all demonstrated the LSTM effectiveness.

An RNN receives information from prior input that may influence the current input and output. The structure of RNN is shown in Figure 1, where the output of previous state h_{t-1} is dependent on the input sequence x_{t-1} with hyperbolic tangent (tanh) as activation function for modelling non-linearity. In theory, a RNN shall have the ability to handle arbitrary long-term dependencies, while in practice, it still suffers from the problem of vanishing and exploding gradients. LSTM is explicitly designed to avoid the gradient problem and the long-term dependency of the data nature. An LSTM has feedback connections composed of three gates as shown in Figure 2, updating and controlling the cell states, which models longer memory that stores and loads information of past events. Because of the feedback connection, an LSTM can process not only single data, but also deal with sequences data. The three gates are the forget gate f_t, input gate i_t and output gate o_t with tanh and sigmoid (σ) activation functions. When processing an LSTM, the first part is called forget gate, a sigmoid function, deciding if the information of the output h_{t-1} of the above unit and the input x_t of this unit should be kept or not from the previous time step. The equation of forget gate can be shown as

$$f_t = \sigma(w_f . [h_{t-1}, x_t] + b_f) \quad (1)$$

The input gate generates a value between [0,1] for each item in $\widetilde{C}_t$, controlling how much new information is added, 0 means completely discarded, 1 means completely pass. The equation of input gate can be shown as

$$i_t = \sigma(w_i \times [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\widetilde{C}_t = \tanh(w_c \times [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \quad (4)$$

After combining the input gate and the cell state, the output gate for prediction is to pass the previous hidden state and the current input into a sigmoid function, then pass the newly modified cell state to the tanh function, which controls the information encoded in the cell state to be the input to the next hidden state. The equation of output gate can be shown as

$$o_t = \sigma(w_o \times [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

The w_f, w_c, w_i, w_o represent the weighting matrix of the hidden state and input, b_f, b_c, b_i, b_o represent the offset value, h is the hidden state, representing short term memory and C is the cell state, representing long term memory.

LSTM controls discarding or adding information through 'doors' to achieve forgetting or memory functions. The forget gate is to decide if the information of the output of the above unit and the input of present unit should be kept or not from the previous time step. The input gate quantifying the importance of the new information in the input and $\widetilde{C}_t$ controls the information to be encoded into the cell state. The output gate is used to control how much of the current cell state is filtered out. The three gates allowed gate layers to have knowledge about the cell state at every instant, which addresses the vanishing gradient problem that makes network training difficult for a long sequence of words or integer. Therefore, LSTM is often used for training long time series data such as weather forecast because of its powerful learning capabilities [22].

3. Input Data Preprocessing

Raw data often come in large set. It is thus necessary to preprocess the data before deep learning. It is also essential to transform the data upon data missing, data unsorted, data scaling, non-stationarity and multicollinearity. In this study, data preprocessing is needed to improve training convergence and reduce training time. The feature in this study, including sea pressure (hPa), temperature (ºC), dew point (ºC), relative humidity (%), wind speed (m/s), wind direction (360 degree), sunshine rate (%), global solar radiation (MJ/m^2), visible mean (km) and cloud amount (0~10), vary so vastly in terms of magnitude. Data preprocessing in z-score normalization is adopted by

$$X' = (X - \mu)/\beta \quad (7)$$

where X is the original data point, X' is the new data point after preprocessing, μ is the mean value of the feature and β is the standard deviation. In aviation, the features are often widely differentiating, for example, the sea pressure is around 1013.25 hPa, but the temperature is in the range of around 9° to 32° in Taiwan. Therefore, it is necessary to scale the data of processing for aviation weather prediction. While training a LSTM model in deep learning, it is often useful to monitor the deep learning training progress.

In this paper, a daily weather data for the whole year of 2020 in Taiwan has been collected from Central Weather Bureau Taiwan. The data contains 10 weather information listed above. Sea pressure was included because of its association with boundary layer inversions, radiative cooling at the surface, clear skies, and the presence of anticyclones. An increase in sea pressure, or persistent high values would provide an indication of a synoptic situation conducive to fog formation. Dew point temperature, the occurrence of fog can be identified by the point at which dry-bulb and dew point temperatures converge. Trends in the relative difference between these two parameters are an indicator of the likelihood of fog formation. These predictors were employed by themselves as well as being combined to calculate relative humidity (an input which was eventually excluded from the final choice of parameters). Relative humidity, while humidity itself is a climate variable, it also affects other climate features. Environmental humidity is affected by winds and by rainfall. Visibility mean, visibility measurements provide short-term nowcasting guidance as the ambient air temperature approaches its dew point and mist begins to form. Total cloud amount, a prerequisite for fog formation is a negative net radiation balance, as inferred by the occurrence of clear nocturnal skies. Sunshine rate, sunshine duration or sunshine rate is a climatological indicator, measuring duration of sunshine in given period (usually, a day or a year) for a given location on Earth, typically expressed as an averaged value over several years. Global solar radiation, the solar radiation that reaches the Earth's surface without being diffused is called direct beam solar radiation. The sum of the diffuse and direct solar radiation is called global solar radiation. Atmospheric conditions can reduce direct beam radiation by 10% on clear, dry days and by 100% during thick, cloudy days. Wind speed, wind speed is a strong determinant of fog. High wind speeds can act to dissipate mist before it forms into a thicker layer of fog. Low wind speeds allow for turbulent mixing, which spreads cooling vertically, deepening the fog layer. Wind direction, dependent on the lead time at which it is sampled, wind direction is either an indicator of synoptic situation (at a long lead) or local conditions (short lead). When employed in unison with wind speed, forecasters can derive an indication of fog likelihood based on the characteristics of both parameters. For example, at the longer lead time, afternoon southeasterlies are usually indicative of a synoptic situation not conducive of fog. However, at a shorter lead time (during nocturnal periods), mild southeasterlies are symptomatic of katabatic drainage flows. While these flows are often necessary for fog formation, they occur almost every night during the year, and as such their presence offers little predictive capability. The above weather information are cross-related. For example, air density is related to air temperature and pressure, so as temperature is also related to wind speed (WS) and wind direction (WD). The data is in a 10 × 366 matrix. Temperature is the feature to be predicted in this study, and the remaining features will be used as input for training the LSTM model.

4. Aviation Weather Prediction

The network by using LSTM is depicted Figure 3. A sequence input layer is first applied to the network in the LSTM layer. By setting with the hidden units of 500 from the hidden state. The hidden state is the output of an LSTM layer, and the hidden unit is corresponding to the amount of information remembered between the hidden state. The fully connected layer is to map output of LSTM layer to a desired output size, and the regression layer is to compute the half-mean-squared-error loss. Before fitting the database into the LSTM, the data will be divided into 90% for training set and 10% for testing set. The former is to fit the model and the latter is to validate the model by comparing its prediction of with the original data.

This paper creates a set of options for training the LSTM network using adaptive moment estimation algorithms [23]. Reduce the learning rate by a factor of 0.005 every epoch. The learning rate is to control how much to change the model in response to the estimated error each time the model weights are updated. If the learning rate is too low, then training takes a long time. If the learning rate is too high, then training might reach a suboptimal or even diverge. The maximum number of epochs for training is set to 250, to measure the number of times of the training sets. The learning rate schedule is set to 'piecewise', which means the software updates the learning rate every certain number of 250 epochs by multiplying with the learning rate factor 0.005. And set learning rate drop factor for 0.2, it is a multiplicative factor to apply to the learning rate every time a certain number of epochs passes. The gradient threshold is set to 1, which is used to clip gradient values that exceed the gradient threshold. After training the model by the training set, a figure of training progress displayed to confirm whether the model is overfitting or not. The testing data is then prepared to make a prediction by the LSTM model. The Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) are employed as performance indicators.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)/y_i\right| \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{9}$$

where n is the number of observed data, y_i is the real value and $\hat{y}_i$ is the forecast value.

The RMSE of using remaining 9 features to predict the temperature is 2.8518, and the MAPE is 14.8993%. In order to obtain better results, each feature is then deleted

and the remaining 8 features are used to train the LSTM model, respectively. The result of each training is shown as Table 1. The purpose is to know which feature has the less effect on predicting the temperature. The results show that when deleted the features of wind speed, wind direction, sunshine rate and global solar radiation has the lowest RMSE than using 9 features for training. For verification that this assumption is correct, the above four features are then using cross-comparison as Table 2 shown below. The first six are deleting 2 features and using the remaining 7 features to train the model, then predict the feature of temperature. The seventh to the tenth are deleting 3 features and using the remaining 6 features to make a training. And the last one is deleting 4 features and using the remaining 5 features to train the model in order to make a prediction. The results of above three training are shown in Figure 4, Figure 5 and Figure 6. Figure 4 is the prediction of using 7 features to train the model, Figure 5 is the prediction of using 6 features to train the model and Figure 6 is the prediction of using 5 features to train the model.

Deleting the feature of wind speed, wind direction, sunshine rate and global solar radiation do decrease the prediction error. Table 3 shows that the lowest 4 RMSE of the combinations by using 8 features are deleting the features of wind speed, wind direction, sunshine rate and global solar radiation, respectively. In order to predict the temperature, at least 6 features are needed for training the model. If the amount of features is reduced to 5, the error will increase to 3.0529, it is larger than the 9 features prediction's RMSE 2.8518, which is used as a benchmark for comparing the rest of the prediction's error. This result shows that the proposed LSTM model can accurately train the parameters and explore the behavior of the deep learning network on the aviation weather numerical data. In this study, a new data point is predicted based on the training of the network, and after using the new data point to retrain the network, another new data point is then predicted. The experimental results are confirmed that the aviation weather numerical data can be well-trained for the proposed LSTM model and make an accurate prediction.

5. Conclusion

This paper presents a weather forecast based on LSTM neural network. By using a dataset of 10×366 matrix, 90% of the data for training and 10% for testing. The results show that different combinations of features will have different effects on weather forecast accuracy. The temperature forecast with the feature combination of sea pressure, dew point, relative humidity, wind speed, global solar radiation, visible mean and cloud amount is the best. It can achieve the lowest RMSE to 2.3872 and decrease the MAPE to 10.8682%.

Data preprocessing is necessary before fitting the data to LSTM in deep learning. It is needed to improve training convergence and reduce training time because the data is usually varying so vastly in terms of magnitude. If there is a lot of irrelevant and redundant information or noisy and unreliable data, it will be more difficult to discover knowledge during the training.

In this paper, an important feature analysis has been developed to modify the model by eliminating redundant attributes. The impact of preprocessing increases the reliability of the long short-term memory network model, which can successfully avoid the complexity of the system; hence the results can increase the reliability of predicting system to further help aviation weather forecasting and strategy planning.

Acknowledgment

This work was supported in part by the Ministry of Science and Technology, Taiwan, ROC under contract MOST 109-2410-H-309-003-.

References

[1] Lewis, F.R., 1922: Weather prediction by numerical process, Cambridge: At the university press.

[2] Charney, J.G., Fjoertoft, R. and Neumann, J.v., 1950: "Numerical integration of the barotropic vorticity equation," Tellus 2(4), 237–254.

[3] Glahn, H.R. and Lowry, D.A., 1972: "The use of model output statistics (MOS) in objective weather forecasting," Journal of Applied Meteorology and Climatology 11(8), 1203-1211.

[4] Schizas, C.N., Michaelides, S., Pattichis, C.S. and Livesay, R.R., 1991: "Artificial neural networks in forecasting minimum temperature (weather)," Artificial Neural Networks 112-114.

[5] Maqsood, I., Khan, M.R., and Abraham, A., 2004: "An ensemble of neural networks for weather forecasting," Neural Computing and Applications 13, 112-122.

[6] Hopfield, J.J., 1982: "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the National Academy of Sciences 79(8), 2554-2558.

[7] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986: "Learning representations by back-propagating errors," Nature 323, 533-536.

[8] Vlachas, P.R., Pathak, J., Hunt, B.R., Sapsis, T.P., Girvan, M., Ott, E. and Koumoutsakos, P., 2020: "Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics," Neural Networks 126, 191-217.

[9] Bocquet, M., Brajard, J., Carrassi, A. and Bertino, L., 2020: "Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization," Foundations of Data Science 2, 55-80.

[10] Geer, A.J., 2021: "Learning earth system models from observations: machine learning or data assimilation," Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences 379(2194), 1-21.

[11] Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadyay, A., Singh, A., Manepalli, A., Chirila,

D., Yu, R., Walters, R., White, B., Xiao, H., Tchelepi, H.A., Marcus, P., Anandkumar, A., Hassanzadeh, P. and Prabhat, 2021: "Physics-informed machine learning: Case studies for weather and climate modelling," Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences 379(2194), 1-36.

[12] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, 2019: "Deep learning and process understanding for data-driven Earth system science," Nature 566, 195-204.

[13] Carrassi, A., Bocquet, M., Bertino, L. and Evensen, G.,2018: "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," WIREs Climate Change 9(5), 1-50.

[14] Hochreiter, S. and Schmidhuber, J., 1997: "Long short-term memory," Neural Computation 9(8), 1735-1780.

[15] Salman, A.G., Heryadi, Y., Abdurahman, E. and Suparta, W., 2018: "Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting," Procedia Computer Science 135, 89-98.

[16] Hong, S.J., Kim, J.H., Choi, D.S. and Baek, K.H, 2021: "Development of surface weather for east model by using LSTM machine learning method," Korean Meteorological Society 31(1), 73-83.

[17] Sangiorgio, M. and Dercole, F., 2020: "Robustness of LSTM neural networks for multi-step forecasting of chaotic time series," Chaos, Solitons and Fractals 139, 110045.

[18] Vlachas, P.R., Byeon, W., Wan, Z.Y., Sapsis, T.P. and Koumoutsakos, P., 2018: "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," The Royal Society A Mathematical Physical and Engineering Sciences 474(2213).

[19] Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H. and Schmidhuber, J., 2009: "A novel connectionist system for unconstrained handwriting recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence 31(5), 855-868.

[20] Gomez, P., Nebot, A., Ribeiro, S., Alquézar, R., Mugica, F. and Wotawa, F., 2003: "Local maximum ozone concentration prediction using soft computing methodologies," Systems Analysis Modelling Simulation 43(8), 1011-1031.

[21] Fente, D.N. and Kumar, S.D., 2018: "Weather forecasting using artificial neural network," Inventive Communication and Computational Technologies 1757-1761.

[22] Li, S., Wang, S.L. and Xie, D.B., 2019: "Research in radar clutter recognition method based on LSTM," The Institution of Engineering and Technology 2019(19), 6247-6251.

[23] Kingma, D. and Ba, J., 2015: "Adam: A method for stochastic optimization," International Conference for Learning Representations, San Diego.

Table 1 The RMSE and MAPE of deleting each feature and using the remaining 8 features for aviation weather forecasting.

Deleted feature	RMSE	MAPE(%)
Sea pressure (SP)	3.0532	16.7177
Dew point (DP)	3.6428	18.0390
Relative humidity (RH)	2.9147	15.0753
Wind speed (WS)	2.5330	11.6201
Wind direction (WD)	2.7290	13.6250
Sunshine rate (SR)	2.4359	11.4736
Global solar radiation (GR)	2.5498	12.3533
Visible mean (VM)	3.5909	17.8495
Cloud amount (CA)	2.8136	13.6726

Table 2 The RMSE of using cross-comparison to delete the feature of wind speed, wind direction, sunshine rate and global solar radiation for aviation weather forecasting.

Predict No.	Deleted feature	RMSE	MAPE(%)
Pred1	WS, WD	2.4998	11.4908
Pred2	WS, SR	2.3872	10.8682
Pred3	WS, GR	2.6472	12.7754
Pred4	WD, SR	2.5108	11.5507
Pred5	WD, GR	2.4035	11.2144
Pred6	SR, GR	2.4160	11.2703
Pred7	WS, WD, SR	2.6265	12.6844
Pred8	WS, WD, GR	2.5727	12.3635
Pred9	WD, SR, GR	2.6655	13.2152
Pred10	WS, SR, GR	2.5396	11.7674
Pred11	WS, WD, SR, GR	3.0529	15.2678

Table 3 The RMSE and MAPE with different feature combinations.

Training features	Feature combinations	RMSE	MAPE (%)
5	SP+DP+RH +VM+CA	3.0529	15.2678
6	SP+DP+RH +WD+VM+CA	2.5396	11.7674
	SP+DP+RH +WS+VM+CA	2.6655	13.2152
	SP+DP+RH +SR+VM+CA	2.5727	12.3635
	SP+DP+RH +GR+VM+CA	2.6265	12.6844
7	SP+DP+RH+WS +WD+VM+CA	2.4160	11.2703
	SP+DP+RH+WS +SR+VM+CA	2.4035	11.2144

	SP+DP+RH+WS +GR+VM+CA	2.5108	11.5507
	SP+DP+RH+WD +SR+VM+CA	2.6472	12.7754
	SP+DP+RH+WD +GR+VM+CA	2.3872	10.8682
	SP+DP+RH+SR +GR+VM+CA	2.4998	11.4908
8	SP+DP+RH+WS +WD+SR+GR+VM	2.8136	13.6726
	SP+DP+RH+WS +WD+SR+GR+CA	3.5909	17.8495
	SP+DP+RH+WS +WD+SR+VM+CA	2.5498	12.3533
	SP+DP+RH+WS +WD+GR+VM+CA	2.4359	11.4736
	SP+DP+RH+WS +SR+GR+VM+CA	2.7290	13.6250
	SP+DP+RH+WD +SR+GR+VM+CA	2.5330	11.6201
	SP+DP+WS+WD +SR+GR+VM+CA	2.9147	15.0753
	SP+RH+WS+WD +SR+GR+VM+CA	3.6428	18.0390
	DP+RH+WS+WD +SR+GR+VM+CA	3.0532	16.7177
9	SP+DP+RH+WS+WD+SR+GR+VM+CA	2.8518	14.8993

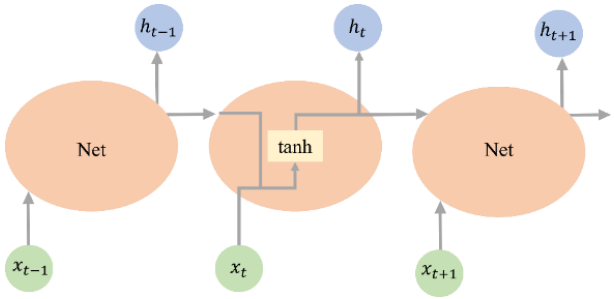

Figure 1 The structure in a standard RNN contains a single layer, where x is the input, h is the output and tanh is the tangent function, the subscript t-1, t, and t+1 is the previous, present, and next time step, respectively.

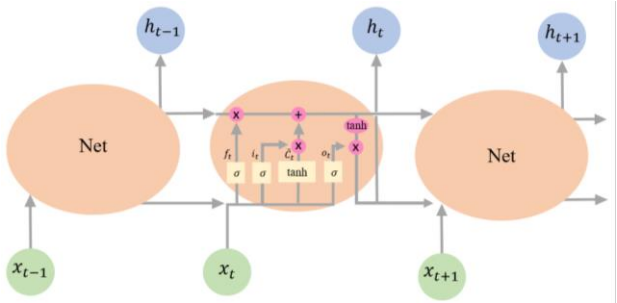

Figure 2 The structure in an LSTM contains four interacting layers by using feedback connection, where $\times$ is the pointwise multiplication, $+$ is the pointwise addition, σ is the sigmoid function, f_t is the forget gate, i_t is the input gate, o_t is the output gate and $\widetilde{C}_t$ is the cell state.

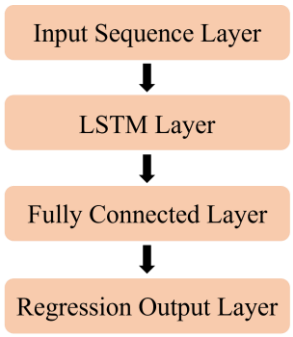

Figure 3 The structure of LSTM model for aviation weather forecasting.

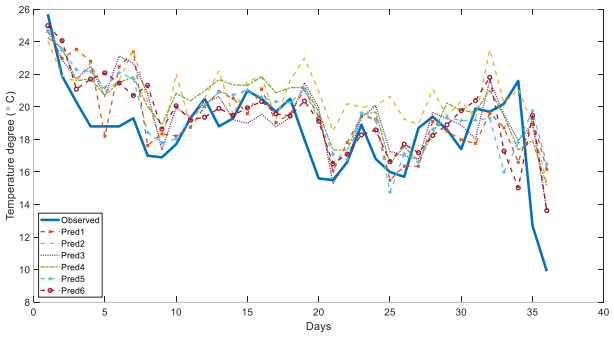

Figure 4 Temperature prediction by using 7 features in Table 3 for aviation weather forecasting.

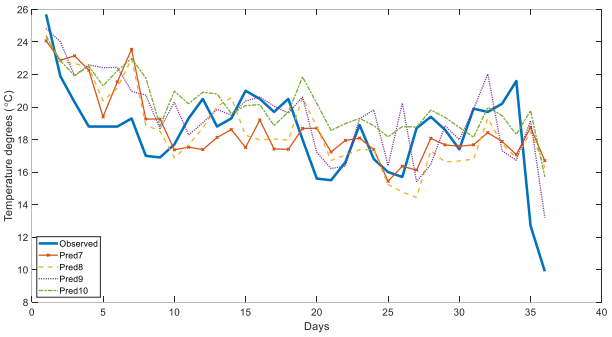

Figure 5 Temperature prediction by using 6 features in Table 3 for aviation weather forecasting.

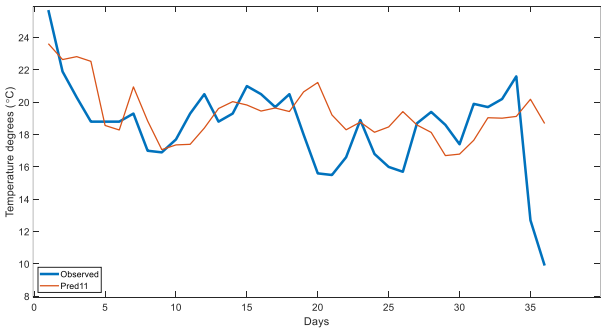

Figure 6 Temperature prediction by using 5 features in Table 3 for aviation weather forecasting.