Scientific
Research
Publishing

The Application of Deep Learning in Airport Visibility Forecast

Lei Zhu[1], Guodong Zhu[2,3], Lei Han[3], Nan Wang[3]

[1]Training Center of Xinjiang Air Traffic Management Bureau, Urumqi, China
[2]College of Atmospheric Science, Nanjing University, Nanjing, China
[3]Meteorological Center of Xinjiang Air Traffic Management Bureau, Urumqi, China
Email: wangnan1986yang@163.com

Abstract

This paper uses Urumqi International Airport's hourly observation from 2007 to 2016 and builds regression prediction model for airport visibility with deep learning method. From the results we can see: the absolute error of hourly visibility is 706 m. When the visibility ≤ 1000 m, the absolute error is 325 m, and this method can predict visibility's trend. So we can use this method to provide the airport visibility's objective forecast guidance products for aviation meteorological services in the future. In this paper, the Urumqi area is as the research object, to explore the depth of learning in the field of weather forecasting applications, providing a new visibility return forecast for weather forecast personnel so as to improve the visibility of the level of visibility to ensure the safe and stable operation of the airport.

Keywords

Deep Learning, Airport Visibility, Regression Prediction

1. Introduction

With the rapid development of the national economy and the increasing popularity of civil aviation transport, airport operation on the visibility is becoming increasingly prominent. A long, low-visibility weather caused by fog, haze and other weather can cause a wide range of airport delays and cancellations. This not only has brought huge losses for the airlines and the airport, but also affects the public travel. At the same time visibility and flight safety are closely related. Low visibility is also one of the most common causes of flight accidents. Urumqi International Airport is the hub of the Xinjiang region airport. It is responsible for the Xinjiang region and Central Asia flight operations. The existing climate data show that Urumqi airport visibility was below 1000 m the average number

of days for 60 days [1] [2]. Most of the low visibility days occurred in the winter half (November to March), up to 57 days. The weather phenomenon that causes low visibility is mainly fog and smoke.

Improving the level of visibility is an important measure to ensure the safe and stable operation of the airport. At present, the low visibility forecast for the smoke, fog and other weather, is still based on empirical forecasts and statistical forecasts. Although with the development of numerical forecasting, there are also numerical and fog model predictions and many experiments have shown that the fog model has only a certain degree of analytical use and is difficult to predict. Therefore, the study of atmospheric visibility is still a difficult and hot spot in meteorological forecast in recent years [3]-[11].

2. Deep Learning

Deep Learning (DNNs) is also known as deep neural network (DNNs), which is the sub-field of machine learning. Its concept originated in the Artificial Neural Network (ANN). In essence, it refers to a class of neural networks with deep structure of the effective training methods. It uses a multi-layer representation to model the complex relationships between the data [12] [13] [14]. Deep Learning can be used in sorting, regression and information retrieval and other specific issues.

3. The Establishment of Prediction Model

3.1. Data Preprocessing

This article uses the Urumqi Airport from 2007 to October 2016 to March the following year 24 hours a day observation data. Contains hourly dominant visibility, temperature, dew point temperature, relative humidity, average wind direction and average wind speed. By sorting and controlling data quality, 43,752 data records were received. Since each factor is composed of different meteorological elements, in order to avoid the difference in magnitude between the various factors, it needs to be normalized before the input factor as the depth neural network, so that its value is limited to [0, 1]. See Equation (1) for the specific algorithm.

$$\text{data} = \left(\text{data} - \min\left(\text{data}\right)\right) \big/ \left(\max\left(\text{data}\right) - \min\left(\text{data}\right)\right) \tag{1}$$

3.2. Prediction of Forecasting Factors

For the time series regression prediction, the simplest way is to build a nonlinear function based on historical data. Combined with the prediction of dominant visibility, we construct two types of factors: the first type of forecasting factor contains only the dominant visibility (Vis) in the past. See Equation (2) for details. Because the dominant visibility is related to the factors such as wind, temperature and relative humidity, the second type of forecasting factor not only includes the dominant visibility in the past, but also the temperature (T), dew point temperature (TD), relative humidity (RH), wind direction (WD) and wind speed (WS). See Equation (3) for details.

$$\text{Vis}_t = f\left(\text{Vis}_{t-1}, \text{Vis}_{t-2}, \cdots, \text{Vis}_{t-n}\right) \tag{2}$$

$$\text{Vis}_t = f\left(\left(\text{Vis}, \text{T}, \text{TD}, \text{RH}, \text{WD}, \text{WS}\right)_{t-1}, \left(\text{Vis}, \text{T}, \text{TD}, \text{RH}, \text{WD}, \text{WS}\right)_{t-2}, \cdots, \\ \left(\text{Vis}, \text{T}, \text{TD}, \text{RH}, \text{WD}, \text{WS}\right)_{t-n}\right) \tag{3}$$

where the dominant visibility is Vis_t for the current dominant visibility. In order to verify the effect of the different length of time on the dominant visibility, here are a number of ns for modeling operations which is used to evaluate the difference in the prediction effect of the model under different time length samples.

3.3. Build a Sample Sequence

This article takes hourly dominant visibility of Urumqi Airport as the forecast object. Using observations from October to March of the following ten years. According to the selection of forecast factors, two kinds of 28 sample sequences were established. See Table 1 for details. In each of the sample sequences, 80% of the 43,752 historical data were randomly selected as training samples for the prediction model training and the predictive effect test. The rest are test samples.

In Table 1, Vis is visibility, T is temperature, TD is Temperature difference, RH is Relative humidity, WD is Wind Direction, WS is Wind speed.

According to the 28 sample sequences which were established the above two types of factor selection methods, the MLP model (Multilayer Perceptron Model) in Keras was used to model the operation. And using test samples to test it. Finally 28 different models of dominant visibility prediction were obtained. The prediction effect of the model is discussed in detail below.

Table 1. The number of two types of forecasting factors at different times.

Number	Value at past n-hour	Number of the first type of forecasting factor	Number of the second type of forecasting factor
		Containing Vis	Containing Vis, T, TD, RH, WD, WS
1	n = 1	1	6
2	n = 3	3	18
3	n = 6	6	36
4	n = 9	9	54
5	n = 12	12	72
6	n = 24	24	144
7	n = 36	36	216
8	n = 48	48	288
9	n = 60	60	360
10	n = 72	72	432
11	n = 84	84	504
12	n = 96	96	576
13	n = 108	108	648
14	n = 120	120	720

4. Predictive Effect Test

This paper constructs the forecasting object for Urumqi airport hourly dominant visibility, which contains a total of nearly 45,000 records. Through the analysis of the forecast object, you can see the Urumqi airport visibility changes in the range of 0 to 10,000 meters. 26.9% of the dominant visibility is 10,000 meters. While the impact of the operation of the civil aviation airport to dominate the visibility of less than 1000 meters accounted for 12.3% of the record. Specific distribution is shown in Figure 1. By using the MLP model to predict the dominant visibility, an hourly dominant visibility prediction result is obtained. Here we examine the predictive effect of this method from two different types of forecasting factors.

4.1. A Model with Dominant Visibility as a Predictor

This type of model contains only the dominant visibility of the past, without adding other meteorological elements. Constructing a predictive model that dominates the dominant time at historical time is based on the dominant visibility at the current time. The prediction model was established by using the dominant visibility of the past 1, 3, 6, 9, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120 hours.

The results of the predictive results shown in Figure 2 show that the average
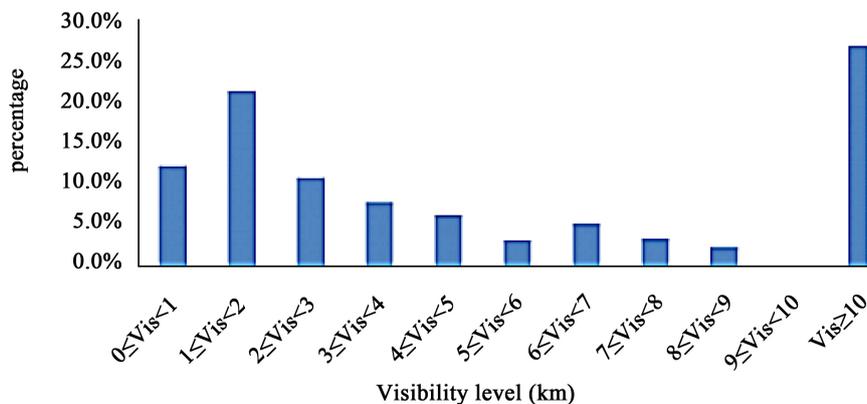

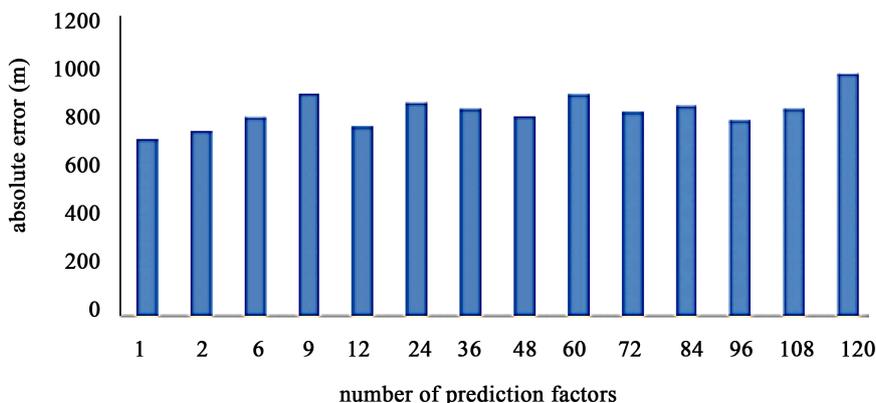Figure 1. Urumqi airport visibility distribution from November to next March.

Figure 2. The results using the dominant visibility of different time lengths.

absolute error of dominant visibility is between 966 m and 706 m using historical dominant visibility as a predictor for different time lengths. It is best to use the forecast for 1 hour before the current time. The average absolute error is 706.98 m. The prediction effect of the model with different time length is different, and the error increases slowly with the increase of the length of time.

In order to fully test the different factors of the model to dominate the visibility, the following statistics by 5000 m within the dominant visibility to predict the average absolute error. And analysing the ability of the model to predict the dominant visibility of different scales. It can be seen from Table 2 that this model has a mean absolute error of less than 1000 m between 325 m and 520 m, and

Table 2. Average absolute error (in m) for visibility below 5000 m.

		0 ≤ Vis < 1000	1000 ≤ Vis < 2000	2000 ≤ Vis < 3000	3000 ≤ Vis < 4000	4000 ≤ Vis < 5000
1	train	327.29	423.56	663.63	840.48	1133.94
	test	326.47	378.32	726.58	897.89	1230.10
3	train	325.97	431.51	690.45	852.26	1134.11
	test	328.46	380.70	741.47	908.79	1208.05
6	train	342.56	446.40	705.54	870.03	1151.82
	test	340.53	398.67	751.12	926.55	1220.44
9	train	456.03	581.39	850.16	991.67	1284.93
	test	428.65	502.76	887.82	1102.37	1337.66
12	train	344.53	452.17	708.21	883.09	1184.77
	test	334.67	406.62	762.61	950.00	1272.34
24	train	380.97	473.22	705.61	865.83	1139.36
	test	367.18	429.12	752.18	938.26	1248.22
36	train	570.81	652.72	887.72	1018.68	1310.13
	test	517.37	591.00	972.87	1170.40	1423.56
48	train	436.89	518.80	754.84	897.60	1173.12
	test	402.15	476.46	821.20	1009.72	1271.40
60	train	437.95	518.15	751.58	893.35	1162.32
	test	408.69	475.60	831.60	996.01	1267.75
72	train	406.06	480.00	707.80	869.73	1137.83
	test	376.18	439.49	775.73	967.95	1244.13
84	train	451.67	517.76	736.56	884.41	1153.88
	test	415.40	471.45	812.99	1003.08	1258.19
96	train	523.64	531.87	735.77	858.77	1119.00
	test	457.69	488.49	819.98	981.49	1234.37
108	train	519.76	542.62	752.20	874.10	1136.56
	test	459.54	503.04	847.17	1017.79	1262.01
120	train	492.18	509.75	711.91	841.43	1092.73
	test	439.76	480.97	793.34	956.49	1200.19

the average absolute error between the training sample and the test sample is small. Which used the past 1, 3 hours two kinds of factors to predict the effect is better. As the length of time in the predictor increasing, the mean absolute error of dominant visibility increases. In addition, as the size of the predicted object increasing, the average absolute error of the dominant visibility in the [4000, 5000] interval is significantly greater than the average absolute error in the range [0, 1000], to about 1200 m.

4.2. A Multi-Meteorological Element Used as a Model of Forecasting Factors

This paper attempts to extend the forecasting factor to the dominant visibility, temperature, dew point temperature, wind direction, and wind direction of the past, due to the fact that there are many reasons for the occurrence of low visibility weather and visibility and temperature, relative humidity and other factors. The model is used to predict the dominant visibility. Through the analysis of the forecast results can be found, using the multi-factor predictor model, the average absolute error in models of different time lengths is predicted from 799 m to 827 m. The average absolute error of the different models is about 10 m. The model used in the past 24-hour multi-factor forecasting factor is best. Its absolute error is 798.87 m. See Figure 3 for details.

In order to fully test the different factors of the model to dominate the visibility, the following statistics the dominant visibility within 5000 m to predict the average absolute error. The ability of the model to predict the dominant visibility of different orders of magnitude. The model has the best predictive effect on the dominant visibility within 1000 m. Its average absolute error is between 450 m and 550 m. Which used the past 72,120 hour factor to build the model to predict the effect is better. With the increase of the dominant visibility level, the error of the model prediction is gradually increased to about 1100 m. The detailed data statistics table is omitted here.
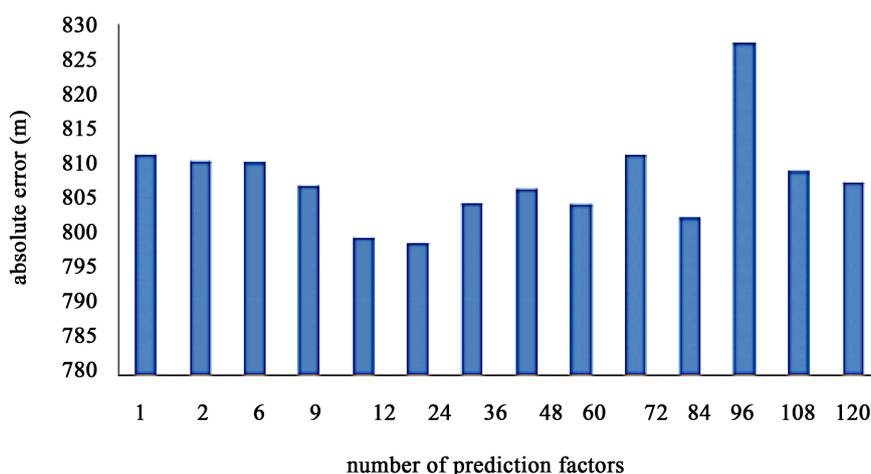

Figure 3. The effect using the dominant visibility of multiple factors for different time lengths.

4.3. Suggestion of Practical Predictive Effect

Based on the model predictive effect constructed by the two kinds of forecasting factors discussed in this paper, the forecasting model of single factor construction is better for the prediction with less than 1000 m visibility. The forecasting model of multi-factor construction is more stable than the forecasting effect of more than 2000 m. So here at the same time use these two models. And taking into account the continuity of the change in visibility, we choose the model using the forecast factor of the past 12 h to make the actual forecast.

Here we select Urumqi Airport December 31, 2016 visibility to predict. The day the airport visibility changes greatly, before 11 hours to maintain more than 1000 m, then quickly decreased to maintain two hours of 100 m, then 16:00 suddenly improved to 2000 m, and then down to maintain at 100 m, see **Figure 4**. The low visibility process includes persistent low visibility, and a sudden improvement in visibility. This has a high test of the ability of the model to predict.

It can be seen from the model predictions that both models can predict the trend of decreasing the visibility and the turn of the day. When the real visibility is greater than 1500 m, the prediction error of the multi-factor model is relatively small. When the dominant visibility is less than 1000 m, the prediction effect of the single factor is relatively better, especially when the long-term continuous visibility is less than 300 m, the average absolute error of the visibility of the single factor is 86 m. Therefore, by using the two types of forecasting models, it is possible to provide a quantitative reference for the forecasting staff to predict the visibility. However, it can be seen from the simulation of the actual case that the depth learning model has a certain hysteresis when the dominant visibility is good or worse, and the error increases obviously when the visibility is greater than 1500 m.

5. Conclusions

Due to the high incidence of low visibility weather, the impact of the system is more complex. Especially the forecasts of low visibility of the starting and ending time are more difficult. So how to as much as possible predict low visibility
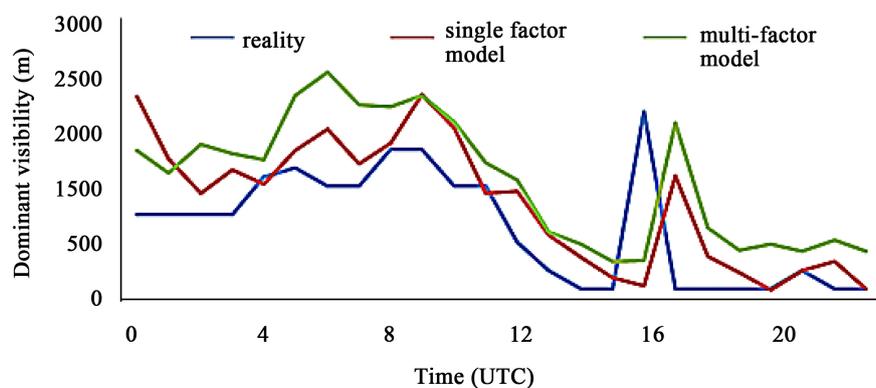

Figure 4. The effect of different models on the forecast of dominant visibility on December 31, 2016.

weather and dissipation of the time are the keys of Urumqi Airport winter service guarantee.

This paper attempts to use the depth neural network for airport visibility prediction, using nearly ten years' data for model prediction. The results of the model forecast show that it can reflect the trend of the dominant visibility of Urumqi Airport. The average absolute error can reach a minimum of 706 m. Where the minimum absolute error of less than 1000 m, it is as low as 325 m. At the same time, we can see that the prediction results of multi-factor factor prediction model are stable. Next, we will try to put this model into the actual business and conduct continuous testing to improve the quantitative forecasting capability of this method in leading visibility.

Although the method has better prediction effect, in the detailed analysis of its forecast results also found some shortcomings, such as predictive visibility turn or turn bad times have a certain lag. The results show that the average absolute error is greater than 2000 m above the dominant visibility, and the prediction effect is less than 1000 m. Then, we will cooperate with low weather conditions, and try to combine the ability to reflect low-level stratification conditions, high-altitude wind field and ground pressure field and other factors as a forecast factor to ensure that the forecasting factor can better contain the low visibility weather conditions to improve the prediction effect of the model.

References

[1] Zhu, L. and Zhu, G.D. (2012) Analysis of Low Visibility Weather Characteristics of Urumqi Airport in Recent 30 Years. *Journal of Civil Aviation Flight Administration of China*, **23**, 27-30.

[2] Zhu, L. and Zhu. G.D. (2010) Application of Support Vector Machine Method in Visual Field Forecast of Airport Runway. *Journal of Stroke*, **29**, 171-175.

[3] Li, P., Wang, S.G., Shang, K.H., *et al.* (2012) Beijing Area Visibility Prediction Based on Neural Network Step-by-Step Classification Modeling. *Journal of Lanzhou University (Natural Science Edition)*, **48**, 52-57. (In Chinese)

[4] Wang, K., Zhao, H., Liu, A.X., *et al.* (2009) Prediction of Atmospheric Visibility Based on Risk Neural Network. *China Environmental Science*, **29**, 1029-1033.

[5] Chen, H.Z., Hao, L.P., Xie, N., *et al.* (2009) Digitalization of Information and Refinement Prediction of Fog. *Journal of Natural Disasters*, **18**, 151-156.

[6] Zhu, G.D. (2011) Multi-Element Prediction of Urumqi International Airport based on SVM Method. Development and Oasis Meteorology, **5**, 40-43.

[7] Peng, S.T., Hu, Y.D., Zhou, R., *et al.* Application of Artificial Neural Network in Urban Short-term Forecast of Visibility. *Proceedings of the* 6*th Annual Conference of Chinese Society for Particles and Cross-Strait Particle Technology Symposium*, 606-609.

[8] Shao, Z.P. (2014) Zhengzhou Airport Visibility Changes and the Causes of Fog Analysis. *Meteorological and Environmental Science*, **37**, 75-82.

[9] He, H. and Luo, H. (2009) Fark Forecasting Method Based on Support Vector Machine Pattern Recognition. *Meteorological Science and Technology*, **37**, 149-151.

[10] Cao, Z.J., Wu, D., Wu, X.J., *et al.* (2008) Climate Characteristics of Fog in China from 1961 to 2005. *Meteorological Science and Technology*, **36**, 556-560.

[11] Li, X.l., Chen, K.J., Wang, K., *et al.* (2008) Classification and Statistical Analysis of Fog in the Capital Airport. *Meteorological Science and Technology*, **36**, 717-723.

[12] Bengio, Y. (2009) Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, **2**, 1-127. https://doi.org/10.1561/2200000006

[13] Hinton, G., Osindero, S. and Teh, Y. (2006) A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, **18**, 1527-1554. https://doi.org/10.1162/neco.2006.18.7.1527

[14] Liu, J.W., Liu, Y., Luo, X.L., *et al.* (2014) Progress in the Study of Depth. *Application Research of Computers*, **31**, 1921-1930, 1942.