Introduction about artificial neural network in precipitation forecast

Zhao Shengrong, (National Meteorological Centre of China, Beijing 100081)

Abstract

In this paper, we used the method of neural network to forecast station precipitation over China. The test result of 2000 indicates that neural network is a feasible method for precipitation forecast. It is better in the light rain forecast, but not good enough for the heavy rain forecast.

1. Introduction about the neural network

The study about neural network has been about 40 years. It has been used in many domains, and obtained great developments. In recent years, it has been used in meteorology widely.

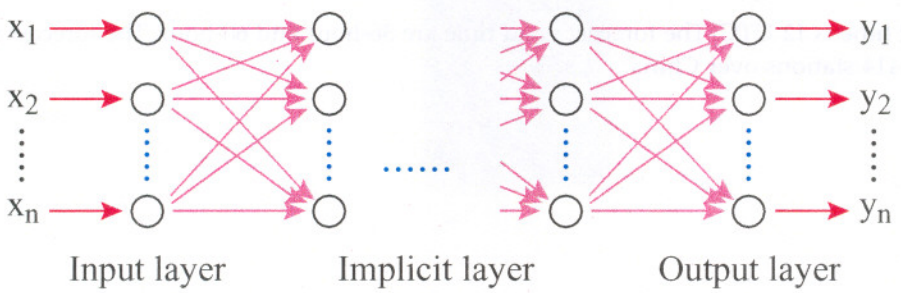
The neural network is a kind of artificial intelligence methods. It was a mathematic mode formed by imitating the process of information deposition and disposal of physical neural network. Nerve cell is the basic structural unit. Single nerve cell is simple and its function is limited, but the neural network composed by a large number of nerve cells is a complicate non-linear system. It can accomplish very complicated action.

The neural network is a non-linear dynamic system. The non-linearity and parallel disposal is the most different point of neural network compared with other methods of NWP products interpretation.

In this paper, we will forecast station precipitation over China by the method of neural network. There are many kinds of artificial neural network modes. What we used is BP (back propagation) network. In the second part, we will introduce the basic structure of BP network and BP algorithm briefly. The third part is about the setting up of forecast equation. In the fourth part, we will test the forecast result from May to September in 2000, and analyse the test result.

2. Basic principles

The BP network is a back propagation network. It contains many layers. It's basic structure is as follows:



The BP network contains input layer, implicit layer and output layer. There can be one or more than one implicit layer on the BP network. The joint on the implicit layer named implicit node. The signals propagate to the implicit nodes from input layer. Then signals of implicit nodes propagate to the next layer after the disposal of weights and operating function. At last, the value on the output nodes are gotten. BP networks can be considered as a non-linear projection from input layer to output layer. We always take Sigmond function as the operating function. In this paper, the operating function is as follows:

$$F(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

Network weights of different nodes are obtained by training. The BP algorithm is used in training process. It's basic process is as follows:

1) Initialisation, that is giving weights initial values.
2) Calculating values of implicit nodes and output nodes, then calculating the error to the observation.
3) Correcting weights through gradient decreasing algorithm according to the error.
4) Repeating the second step and the third step, until the network convergent or the error is in the limit given in advance.
5) Getting the file of weights of nodes and the file of network parameters. At last, the training ending.

3. Forecast model

First, interpolate T106 output on grid to stations during May to September from 1995 to 1998. Then, calculate the correlation coefficients between the observation precipitation and the forecast factors of same station. After that, select factors that absolute value of correlation coefficient is large enough as the last forecast factors, at the same time, considering the representative ability and the NWP precision of the factor. Then form the data file and begin the network training. At last got weight files and network parameter files of different station and different valid forecast time. Then forecast equations are founded.

The BP network in this paper contains one implicit layer and there are 20 implicit nodes on implicit layer. There are three output nodes, corresponding three different situations of precipitation, respectively. The first node represents the situation without precipitation. The second node represents the situation that precipitation is more than 0mm but less than 25mm. The third node represents the situation that precipitation is equal to or more than 25mm. The disposal in detail is as follows:

1) If the precipitation is equal to 0 mm, the first node is 1, but the second and third node are 0;
2) If the precipitation is more than 0 mm but less than 25 mm, the second node is 1, but the first and third nodes are equal to 0;
3) If the precipitation is equal to or more than 25 mm, the third node is 1, but the first and second nodes are 0.

By using the data during May to September from 1995 to 1998, forecast equations of 414 stations over China are established. The forecast result is values of three output nodes. Select the node with maximum output value as the precipitation forecast. That is:

1) If the value of the first output node is maximum, then the precipitation forecast is without precipitation;
2) If the value of the second node is maximum, then the precipitation forecast is more than 0 mm but less than 25 mm;
3) If the value of the third node is maximum, then the precipitation forecast is equal to, or more than, 25 mm.

4. Information about products

The forecast start time is 12 UTC. The forecast valid time are 36-hour and 60-hour. The forecast results are precipitation grades of 414 stations over China.

5. Testing of forecast results

Test results of precipitation forecast from May to September in 2000 are shown in the following tables:

Test result for light rain

	valid time	Ts score	Leak forecast rate	False alarm rate
May	36-hour	0.38	0.50	0.39
	60-hour	0.31	0.58	0.46
June	36-hour	0.45	0.44	0.30
	60-hour	0.38	0.51	0.30
July	36-hour	0.38	0.49	0.30
	60-hour	0.33	0.54	0.45
August	36-hour	0.45	0.30	0.45
	60-hour	0.37	0.52	0.38
September	36-hour	0.41	0.46	0.39
	60-hour	0.34	0.54	0.43

Test result for heavy rain:

	Valid time	Ts score	Leak forecast rate	False alarm rate
May	36-hour	0.02	0.98	0.43
	60-hour	0.00	1.00	0.17
June	36-hour	0.05	0.94	0.55
	60-hour	0.01	0.99	0.20
July	36-hour	0.03	0.97	0.62
	60-hour	0.01	0.99	0.23
August	36-hour	0.04	0.96	0.68
	60-hour	0.01	0.99	0.14
September	36-hour	0.05	0.84	0.46
	60-hour	0.00	1.00	0.09

From the first table, we can conclude that this method is better for light rain forecast. The 36-hour cumulative Ts score of light rain forecast in those four months are 0.38, 0.45, 0.38, 0.45, 0.41 respectively. It reaches 0.45 in June and August. Leak forecast rates and the absent forecast rates are equal roughly. The corresponding Ts score for 60-hour forecast is 0.31, 0.38, 0.33, 0.37, 0.34 respectively. comparing with 36-hour forecast, leak forecast rates of 60-hour forecast increase, and leak forecast rates are lager than absent forecast rates. It also can be seen that the forecast for heavy rain is not good enough. The cumulative Ts score of heavy rain for 36-hour forecast is 0.05 in June, and it is 0.01 for 60-hour forecast. In conclusion, the leak forecast of heavy rain is prominent.

From the analysis before, it can be indicated that the neural network is a feasible method for the precipitation forecast. It is better in the light rain forecast, but not good enough for the heavy rain forecast. In an aspect, the forecast ability of this method for the small probability events needs to be discussed more deeply. In other aspect, more researching works are needed for disposal to the observation precipitation and forecast factors. For this purpose, we are researching other method to improve the forecast of heavy rain.