**Appendix:**

**1 Long Short-Term Memory Neural Network**

In 1997, the LSTM model was first proposed by Hochreiter and Schmidhuber [7]. LSTM is a time recursive neural network, specifically designed to solve the long-term dependency problem of general Recurrent Neural Network (RNN).

All recurrent neural networks have the form of a chain of repeated modules of the neural network. In a standard RNN, the repeating module will have a very simple structure, such as a single tanh layer.

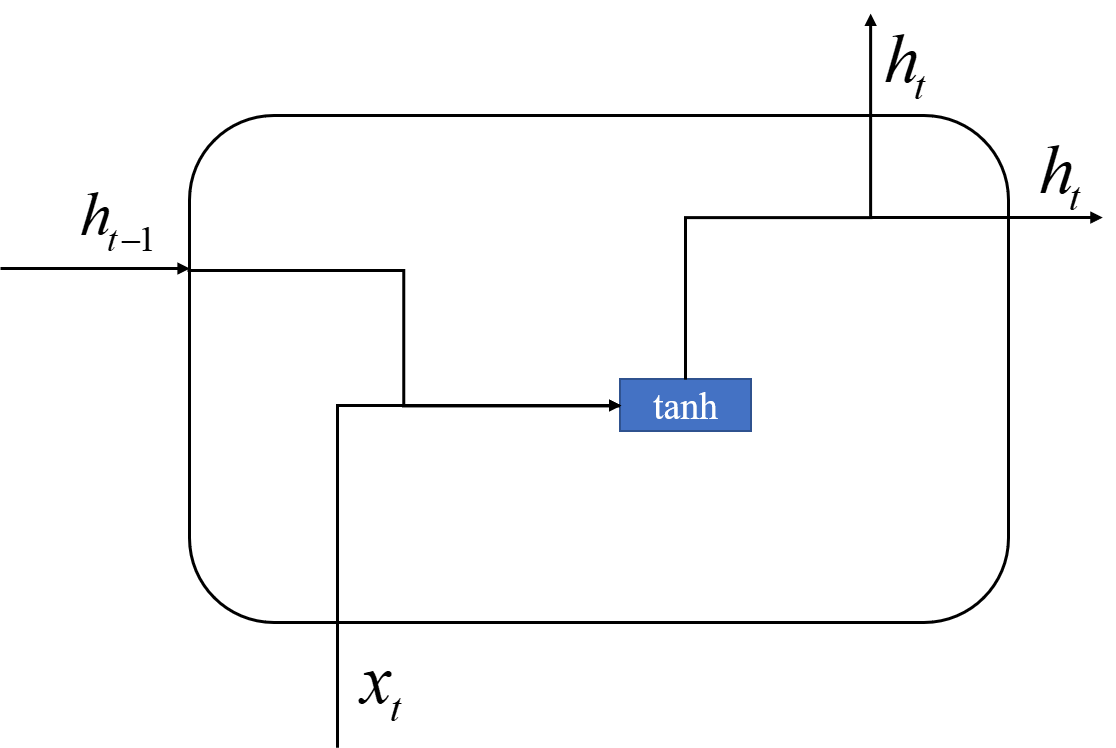


Fig 1. RNN model structure

LSTM also has this chain structure, but the repeating module has a different structure.

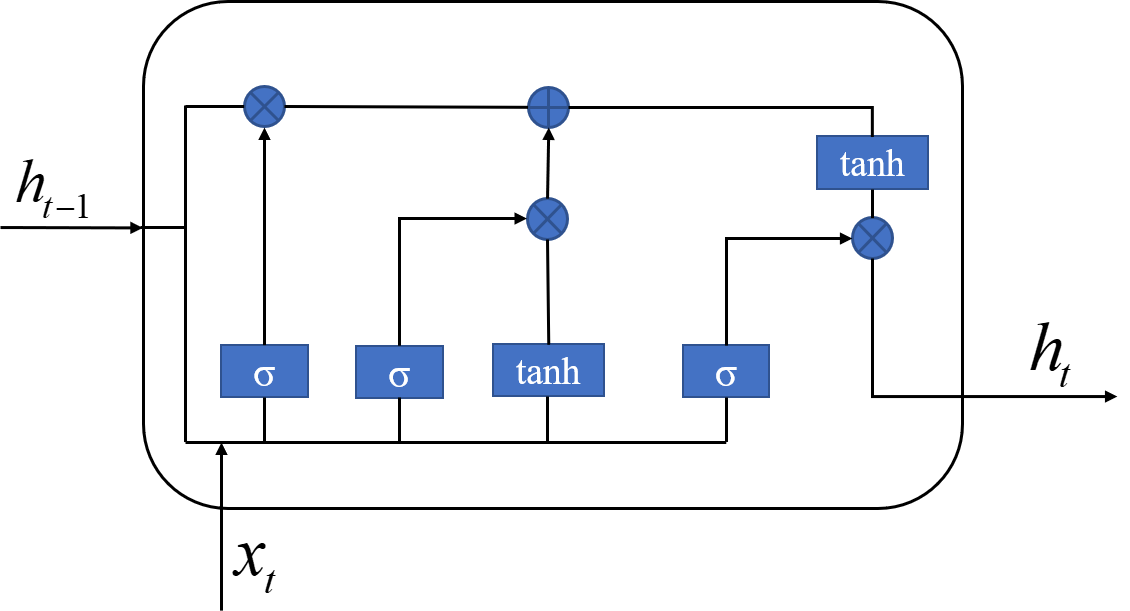


Fig 2. LSTM model structure

The first step of LSTM is to decide what information we want to discard from the cell state. It looks at  (previous output) and  (current input), and outputs a number between 0 and 1 for each digit in the cell state  (previous state). 1 represents completely retained, and 0 represents completely deleted. The following Eq.(1) represents the information  retained after passing through the forgotten door. Where  represents the output of the previous cell,  represents the output of the current cell, and  is the sigmoid function.

 (1)

The sigmoid layer called the “input gate layer” determines which values we will update. The next tanh layer creates a candidate vector , which will be added to the state of the cell. In the next step, combine these two vectors to create an update value.  s memory gate value, and  represents temporary cell layer.

 (2)

 (3)

Now it’s time to update the last state value  to . Multiply the last state value  by  to express the part that is expected to be forgotten. Then we will get value plus . The resulting value is the new candidate, according to how much we have decided to update the value of each state is measured.

 (4)

Finally, we need to decide what we want to output. First, we run a sigmoid layer, which determines what part of the state of our cells to be output. Then, we pass the cell state through tanh, normalize the value between -1 and 1, and multiply it by the output of the sigmoid gate, so we only output those parts that we decided.

 (5)

 (6)

In the above formulas, ,  and  are the weight of each layer, in addition, ,  and  represents the offsets of each threshold layer.

**2 Bi-directional Long Short-term Memory Neural Network**

BILSTM [34] is the abbreviation of Bi-directional Long Short-Term Memory, which is a combination of forward LSTM and backward LSTM.

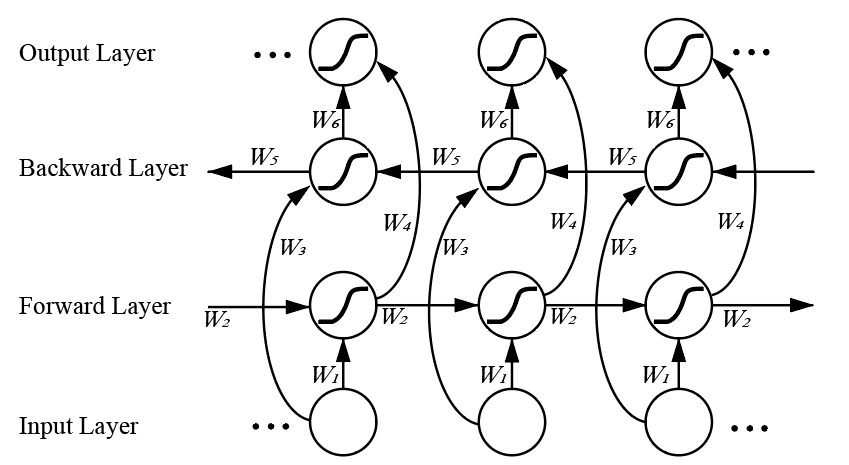


Fig 3. BILSTM model structure

In the forward layer, the forward calculation is performed from time 1 to time  , then the output of each forward hidden layer is saved. In the backward layer, calculate backward along the time  to time 1, and get the output of each backward hidden layer and save it. Finally, at every moment of  corresponding to the results of the forward layer and the backward layer to obtain the final combined output. The mathematical expression of BILSTM model is as follows:

 (7)

 (8)

 (9)

In the Eqs.(7)-(9),  and  represent the output of forward hidden layer and backward hidden layer respectively.  represents the final output. Additionally, , , , ,  and  represent the weight of each calculation layer.