Research on recurrent neural network model based on weight activity evaluation

Cheng Zhang[1], *Luying* Li[1], *Yanmei* Liu[1], *Xuejiao* Luo[1], *Shangguan* Song[1], and *Dingchun* Xia[1,2,*]

[1]School of Information Engineering, Wuhan Institute of Design and Sciences, Wuhan, Hubei, China
[2]School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan, Hubei, China

Abstract. Given the complex structure and parameter redundancy of recurrent neural networks such as LSTM, related research and analysis on the structure of recurrent neural networks have been done. To improve the structural rationality of the recurrent neural network and reduce the amount of calculation of network parameters, a weight activity evaluation algorithm is proposed that evaluates the activity of the basic unit of the network. Through experiments and tests on arrhythmia data, the differences in the weight activity of the LSTM network and the change characteristics of weights and gradients are analyzed. The experimental results show that this algorithm can better optimize the recurrent neural network structure and reduce the redundancy of network parameters.

Keywords: Weight activity evaluation, Parameter redundancy, Recurrent neural network.

1 Introduction

A recurrent neural network is a dynamic artificial neural network that has outstanding performance in dealing with time and space correlation problems and can better capture useful information in time series. The cell of LSTM [1] consists of three complex gate structures that control the flow of information. Compared with the original RNN, the dynamic performance of LSTM has been improved to a certain extent. Although LSTM can capture information in long-term sequences, there are still difficulties. Travis proposed [2] an evolutionary approach to implementing a sparse recurrent neural network to reduce network parameters. Although the goal of lowering network parameters is finally achieved, the training time of this method is quite long, and the efficiency is low. The RNN time step often depends on practical problems and the experience of researchers. Zahra proposed a time-step self-organized RNN [3]. It still takes a lot of time to find the optimal time step. Alexander Ororbia introduced various cell structures such as RNN, GRU[4], LSTM, MGU [5], and UGRNN[6]. He proposed an evolutionary method to dynamically combine various cells to generate a hybrid recurrent neural network [7], a sparse and non-fully linked

* Corresponding author: kristoffzc@163.com

network, compared to the fully connected recurrent neural network of a single Cell. ElSaid[2] pointed out that when the number of LSTM network weights was reduced by 42% to 45%, the accuracy rate of the neural network did not drop, indicating that there are partially redundant weights in the LSTM network.

This paper focuses on the research and analysis of the recurrent neural network structure and the neurons activity efficiency, proposes a weight activity evaluation algorithm (WAEA), and establishes a dynamic sparse recurrent neural network based on the WAEA. WAEA evaluates the activity of weights according to the correlation between weights and gradients, selects the weights with less activity, discards the weights with less activity, and retains the weights with greater activity to simplify the structure and parameters of the network. Experimental analysis shows that the algorithm optimizes the recurrent neural network structure, improves the dynamic network performance, and alleviates the problem of parameter redundancy.

2 Methodology

2.1 Recurrent neural network

Origin recurrent neural network (RNN) has a relatively simple recurrent structure without a complex gate structure, as shown in equation (1). Figure 1 sketches a three-layer recurrent neural network.

$$\begin{aligned} S_t &= f(U \cdot X_t + W \cdot S_{t-1}) \\ O &= g(V \cdot S_t) \end{aligned} \tag{1}$$

As shown in table 1, it can be seen that the original RNN is significantly different in the number of parameters from the more complex LSTM, GRU [4], and MGU [5]. Assume a three-layer network, whose dimensions are m, n, and q, respectively, as shown in figure 2. U is the weight between the output layer and the hidden layer. V is the weight between the hidden layer and the output layer. The hidden layer can be the original RNN, LSTM, GRU, or MGU. The dimensions of weight U and W of LSTM are four times that of the original RNN, while GRU and MGU are three times and twice that of the original RNN, respectively. The increase in the number of parameters is precisely due to the gate. As a result, the gate increases the number of model parameters to a certain extent, which may take more time during the model training process.

Table 1. The parametric scale of recurrent neural networks.

Model	Number of gates	U	V	W
RNN	0	$\mathrm{m} \times n$	$\mathrm{n} \times q$	$\mathrm{n} \times n$
LSTM	3	$4 \times \mathrm{m} \times n$	$\mathrm{n} \times q$	$4 \times \mathrm{n} \times n$
GRU	2	$3 \times \mathrm{m} \times n$	$\mathrm{n} \times q$	$3 \times \mathrm{n} \times n$
MGU	1	$2 \times \mathrm{m} \times n$	$\mathrm{n} \times q$	$2 \times \mathrm{n} \times n$

2.2 Weight activity

ElSaid [2] figured out that some weights are not crucial to the network's performance. These weights have less impact on the performance of the network. In this paper, the

weights that have no or negligible influence on the network performance are called silent weights. On the contrary, the weights that play a vital role in the network's performance are called active weights. In figure 2, the solid line represents the more active weight, and its activity is relatively high. In contrast, the dotted line represents the less active weight, and its activity is relatively low.

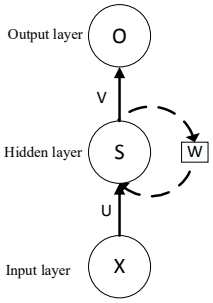


Fig. 1. Recurrent neural network.

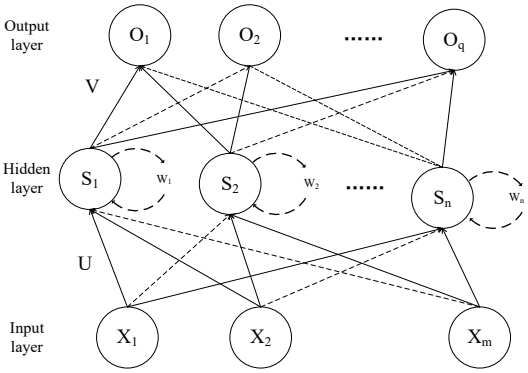


Fig. 2. Multilayer recurrent neural network.

2.3 Weight activity evaluation algorithm

The weight activity evaluation algorithm (WAEA) mainly evaluates the activity of the weights, as shown in algorithm 1. Compared with the active weights, the silent weights have little impact on the performance, so some weights with less activity can be appropriately removed. WAEA measures the activity of weights mainly based on two indicators:

(1) Weight. The initial value of the weight has a significant influence on the network. If the weight is too large, the gradient will disappear. If the weight is too small, the gradient will explode. The weight is an indicator for evaluating the activity of the weight.

(2) Gradient. The weights with a significant gradient trend are in the continuous learning process during network training. The weights with a gentle movement are relatively saturated, and the gradient disappears, so the gradient trend is an index to evaluate the weight activity.

The *Filter* is the gaussian filter function in equation (2), which is used to filter the abnormal weight. After filtering, the minimum and maximum weights will be filtered out. $Activity$ is the activity function that evaluates the activity of the weights. Suppose a multilayer recurrent neural network is trained n times over a certain period. w_i is the ith

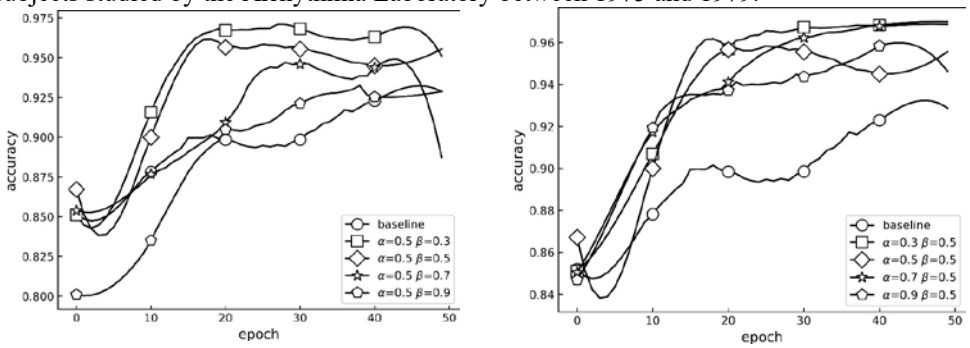
network weight, g_i is the gradient of the i th weight, I is the variance of the gradient during n times of network training, and Q is the mean of the weight during n times of network training.

$$
\begin{aligned}
&Filter(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \\
&I = \frac{\sum_{i=1}^{n}(g_i - \frac{\sum_{j=1}^{n} g_j}{n})^2}{\mathrm{n}} \\
&Q = \frac{\sum_{i=1}^{n} w_i}{\mathrm{n}} \\
&Activity = \alpha \cdot Filter(Q) + (1-\alpha)\cdot I
\end{aligned}
\tag{2}
$$

The activity function comprehensively considers the weight means and gradient variance, and its value is determined by the filtered weight to mean and gradient variance. The first part of equation (2) comes from the mean of the weight, which measures the overall level of the weight value in a period. When the weight is generally tiny or large, the filtered weight significantly affects the activity. The second part comes from the variance of the gradient. When the gradient change is usually gentle, the contribution of the variance of the gradient to the activity is also more negligible. Equation (2) is essential for the weight activity evaluation algorithm. According to the weight activity, some weights with low activity are selectively discarded, and weights with high activity are retained to simplify the neural network structure.

3 Analysis

The experiment uses the MIT-BIH arrhythmia dataset. MIT-BIH is an arrhythmia database provided by the Massachusetts Institute of Technology. MIT-BIH is one of the internationally recognized standard ECG databases. In recent years, it has been widely used in related fields. The dataset contains 48 half-hour two-channel Holter recordings from 47 subjects studied by the Arrhythmia Laboratory between 1975 and 1979.



(a) $\alpha = 0.5$ (b) $\beta = 0.5$

Fig. 3. Accuracy.

Figure 3 compares the accuracy of the original LSTM network and the LSTM network using the WAEA. The baseline is the accuracy of the LSTM network on the test set. In

figure 3(b), when the α are 0.3, 0.5, 0.7, and 0.9, respectively, and β is 0.5, the accuracy of the sparse LSTM network generated by WAEA is higher than the original LSTM network. When α is 0.3, the accuracy is the highest. When α is 0.5 or 0.9, accuracy is relatively poor. As illustrated in figure 3(a), when the weight of 50% is discarded, the accuracy of the sparse network remains at about 95%. However, when the ratio of discarded weights rises to more than 70%, the accuracy of the sparse network declines. When 90% weight is discarded, the performance of the sparse network is reduced compared with the original LSTM. There are a lot of redundant weights in neural networks, and these weights will affect the speed and accuracy of network training. And properly removing these redundant weights can optimize the structure of the recurrent neural network and improve the accuracy of the network to a certain extent.

Algorithm 1 WAEA

Data: Data set $X, x_i \in X$;Label $Y, y_i \in Y$;Parameter α

Output: Activity

begin:
initialize net Net;
initialize set W;
initialize set G:
for $x_i \in X$, $y_i \in Y$ **do**
pred = Net(x_i);
compute Loss(y_i , pred);
g_i = compute gradient;
w_i = update weight of Net;
append g_i to G;
append w_i to W;
end

$$Filter(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2});$$

$$I = \frac{\sum_{i=1}^{n}(g_i - \frac{\sum_{j=1}^{n} g_j}{n})^2}{\mathrm{n}};$$

$$Q = \frac{\sum_{i=1}^{n} w_i}{\mathrm{n}};$$

$$Activity = \alpha \cdot Filter(Q) + (1-\alpha)\cdot I;$$

return Activity;
end

4 Conclusion

Aiming at the parameter redundancy problem in the training of the recurrent neural network, this paper analyzes the characteristics of the weight and gradient of the recurrent neural network in the training process and the difference of the weight activity. On this basis, a weight activity evaluation algorithm (WAEA) is proposed. WAEA quantifies the activity of weights and provides a reference method for measuring the pros and cons of weights. The experimental results show that WAEA has a better performance in measuring the weight

activity of LSTM, and it is found that there is an overall difference in the weight activity of each gate structure of LSTM, mainly because the weight activity of the input gate is generally lower than that of other gate branches. The weight activity of the cell is usually higher than that of other gate branches. The spatial structure storage efficiency of the LSTM network optimized by WAEA is improved, and the computing efficiency and accuracy are also improved.

This work was supported by the Scientific Research Project of the Education Department of Hubei Province under grant B2021377.

References

1. HOCHREITER, SCHMIDHUBER. Long short-term memory [J].Neural computation, 1997, 9(8):1735-80.
2. ELSAID A, JAMIY F E, HIGGINS J, et al. Optimizing long short-term memory recurrent neural networks using ant colony optimization to predict turbine engine vibration [J].Applied Soft Computing, 2018, 73:969 - 91.
3. ABBASVANDI Z, NASRABADI A M.A self-organized recurrent neural network for estimating the effective connectivity and its application to EEG data [J].Computers in Biology and Medicine, 2019, 110:93 - 107.
4. CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB]. arXiv:14123555,2014.
5. ZHOU G-B, WU J, ZHANG C-L, et al. Minimal gated unit for recurrent neural networks [J].International Journal of Automation, 2016, 13(3):226-34.
6. COLLINS J, SOHL-DICKSTEIN J, SUSSILLO D. Capacity and trainability in recurrent neural networks [EB]. arXiv:161109913,2016.
7. ORORBIA A, ELSAID A, DESELL T. Investigating recurrent neural network memory structures using neuro-evolution [C]//Proceedings of the Genetic and Evolutionary Computation Conference. ACM,2019:446-55.