Prediction of Sea Clutter Based on Recurrent Neural Network

Wenyuan Wang¹, Tingting Ji² and Jinghao Sun¹

¹College of Information Science and Engineering, Ocean University of China, Qingdao, Shandong, China
²Teaching Center of Fundamental Courses, Ocean University of China, Qingdao, Shandong, China

E-mails: wangwenyuan0810@163.com, jtt@ouc.edu.cn

Abstract. As radar sea clutter has chaotic characteristics, a prediction model can be established to predict sea clutter according to short-term predictability of chaotic time series. Therefore, this paper proposes a sea clutter amplitude prediction model based on LSTM, Stacked LSTM and Nested LSTM. This paper first obtains the best state space for sea clutter prediction according to the phase space reconstruction theory and establishes the prediction equation, then uses three networks to predict the sea clutter amplitude, and finally compares the effects of the three neural networks and analyzes the factors that affect the prediction precision. The experimental results on the IPIX radar data and P-band radar data show that the model used in this paper has a smaller prediction error than RBF, and the prediction performance is better. The model can achieve high-precision and high-efficiency prediction of sea clutter and has verified the effectiveness of the method.

1. Introduction

Sea clutter is the backscattering echo from radar illuminating the sea surface. The analysis and research on the characteristics of sea clutter has not only theoretical significance, but also practical importance in applications such as sea surface target detection. For a long period of time, sea clutter is considered to be a completely random signal. Therefore, the research and processing of sea clutter mostly start from statistical theory. Researchers have proposed many statistical distribution models, such as Rayleigh distribution, log-normal distribution, Weibull distribution, compound K distribution. But these statistical models have limited application scope and cannot reflect the inherent sea clutter mechanism well.

In the 1990s, Simon Haykin and others repeatedly studied the sea clutter data, and confirmed chaotic characteristics of sea clutter. Meanwhile the chaotic time series has short-term predictability. Therefore, a prediction model can be established to study the internal nonlinear dynamics of sea clutter[1,2]. The research shows that the neural network can be used in the short-term prediction of sea clutter. In[3], Leung et al. used radial basis function (RBF) neural network to predict sea clutter, and used prediction error to detect sea ice. Subsequently, the team proposed to use multiple RBF neural networks to predict sea clutter, and experimental result showed that the effectiveness of this method is better than that of a single model[4]. Lu Ning et al. constructed a Generalized Regression Network (GRNN) target detector in[5], and completed target detection in sea clutter by using the difference in prediction errors between sea clutter and target echo. In 2015, Xu Ting proposed a sea clutter prediction method based on Genetic Wavelet Neural Network, which used IPIX radar data for multi-step prediction, and achieved better prediction effect than general neural network[6]. Zhao et al. proposed a method for predicting the near sea clutter power based on long short-term memory network, which achieved high prediction
Ma Liwen et al. predicted the amplitude of the measured sea clutter data using the gated cyclic network (GRU), and achieved good prediction results[8].

In this paper, three kinds of recurrent neural networks are used to predict the sea clutter amplitude, which are LSTM[9], stacked LSTM (stacked LSTM) and nested LSTM (nested LSTM) [10]. And then performance of the three networks in sea clutter prediction is analyzed. Finally the paper analyze the performance of the three networks in sea clutter prediction.

2. Nonlinear principle of sea clutter prediction
For chaotic time series, the establishment and prediction of chaotic model are carried out in phase space. Thus the reconstruction of phase space is a very important step for the processing of chaotic time series.

2.1. Phase space reconstruction
The basic idea of phase space reconstruction is: the evolution of any component in the system is determined by other components interacting with it. Therefore the information of these related components is implicit in the development of any component.

Takens proposed the embedding theorem in 1981. For an infinite one-dimensional scalar time series \(\{x_i, 1 \leq i \leq n\}\) of \(d\)-dimensional chaotic attractor without noise, an \(m\)-dimensional embedded phase space can be found in the sense of topological invariance, as long as the dimension is \(m \geq 2d + 1\). According to Takens embedding theorem, a phase space in topological sense can be reconstructed from one-dimensional chaotic time series. There are two methods for phase space reconstruction: derivative reconstruction and coordinate delay reconstruction. The second method is used to reconstruct the phase space in this paper. For chaotic time series \(\{x_i, 1 \leq i \leq n\}\), the expression of phase space reconstruction is shown in equation (1):

\[
y_i = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\} \quad i = 1, 2, \ldots, M
\]

Where \(y\) is the point in the phase space, \(\tau\) is the delay time, \(m\) is the embedding dimension, and \(M\) is the number of points in the phase space, \(M = n - (m-1)\tau\). Therefore, the delay time \(\tau\) and embedding dimension \(m\) need to be determined for the phase space reconstruction of chaotic time series.

2.1.1. Delay time \(\tau\). If the delay time \(\tau\) is too small, the values of the two adjacent coordinate components in the phase space will be relatively close. Thus, two independent components cannot be obtained. Contrarily, the two adjacent coordinate components will be completely independent, and the projection of chaotic attractor trajectory in two directions will lose correlation. Therefore, it is necessary to choose a suitable \(\tau\) to achieve a balance between independence and correlation. The commonly used methods to calculate the delay time include autocorrelation method, average displacement method, mutual information method and so on. In this paper, we use the mutual information method to calculate the delay time.

If the discrete time series is \(\{x_i, 1 \leq i \leq n\}\) and \(P_i(x_i)\) is the probability density of point \(\{x_i\}\), then the information entropy of the sequence \(\{x_i\}\) is:

\[
H(X) = -\sum_{i=1}^{N} P_i(x_i) \log_2 P_i(x_i)
\]

Let \(y_i = x_{i+\tau}\), that is, the sequence \(\{y_i\}\) is a delay sequence of sequence \(\{x_i\}\). Then the conditional entropy of \(\{y_i\}\) for \(\{x_i\}\) is:

\[
H(Y|X) = -\sum_{i,j=1}^{N} P_{xy}(x_i, y_i) \log_2 \frac{P_{xy}(x_i, y_i)}{P_x(x_i)} = H(X,Y) - H(X)
\]

Where \(P_{xy}(x_i, y_i)\) is the joint probability density of the points \(x_i\) and \(y_i\), and the joint entropy of the variables \(X\) and \(Y\) is defined as:
\[
H(X,Y) = -\sum_{i,j} P_{xy}(x_i, y_j) \log_2 P_{xy}(x_i, y_j)
\]

From this, the mutual information of \(X\) and \(Y\) is:
\[
I(Y, X) = H(Y) - H(Y|X) = H(Y) + H(X) - H(X,Y)
\]

\(I(Y, X)\) is a function related to delay time \(\tau\). In order to simplify calculation, \(\tau\) corresponding to the minimum value of \(I(Y, X)\) is selected as the optimal delay time.

### 2.1.2. Minimum embedding dimension \(m\).

The embedding dimension is the minimum dimension to restore the chaotic dynamic system. If the \(m\) value is too small, the geometric structure of the chaotic system cannot be completely expressed. If the \(m\) value is too large, the geometric structure of the original system can be completely expressed in theory. But at the same time, the calculation amount will be increased and the noise will be amplified. For the determination of the minimum embedding dimension, the commonly used methods include geometric invariant, false nearest neighbors and Cao method. In this paper, Cao method is used. In this method, only the delay time \(\tau\) is needed and the embedding dimension \(m\) can be obtained with a small amount of data.

For time series \(\{x_i, 1 \leq i \leq n\}\), when the delay time is \(\tau\) and the embedding dimension is \(m\), the \(Y_i(m)\) is \(Y_i(m) = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}\) and its nearest neighbor \(Y_{NN}(m)\) is \(Y_{NN}(m) = \{x_{iNN}, x_{i+\tauNN}, \ldots, x_{i+(m-1)\tauNN}\}\). Three variables are be defined with:

\[
a_2(i,m) = \frac{|Y_i(m+1) - Y_{iNN}(m+1)|}{|Y_i(m) - Y_{iNN}(m)|}
\]

\[
E(m) = \frac{1}{N-m\tau} \sum_{i=1}^{N-m\tau} a_2(i,m)
\]

\[
E_i(m) = \frac{E(m)}{E(m+1)}
\]

For a certain time series, \(E_i(m)\) will not change after \(m\) is greater than a certain fixed value \(m_0\), and the value of \(m_0\) can be used as the best embedding dimension of phase space reconstruction. However, for random sequences, with the increase of \(m\), \(E_i(m)\) not reaching a stable value, so it is difficult to distinguish whether \(E_i(m)\) is slowly changing or converging. Thus a definition is added:

\[
E^*(m) = \frac{1}{N-m\tau} \sum_{i=1}^{N-m\tau} |x_{i+m\tau} - x_{iNN}|
\]

\[
E_2(m) = \frac{E^*(m+1)}{E^*(m)}
\]

If there is no correlation between the data, \(E_2(m)\) is always equal to 1. For chaotic time series, the correlation between data depends on the value of \(m\). Thus there always exists \(m\) so that \(E_2(m)\) is not equal to 1. According to the curves of \(E_i(m)\) and \(E_2(m)\), we can find the minimum embedding dimension \(m\) when \(E_i(m)\) tends to be stable and \(E_2(m)\) is not equal to 1.

### 2.1.3. Sea clutter prediction equation.

If the sea clutter sequence is \(\{x_i, 1 \leq i \leq n\}\), the expression of the sea clutter phase space reconstruction is as follows:

\[
x_{i+m\tau} = F\{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}
\]
The above equation is the sea clutter prediction equation. If the expression of the state equation $F$ can be obtained, the sea clutter prediction can be carried out through this equation. In order to increase the amount of information in the system and enhance the prediction equation in function generalization, this paper actually uses the prediction equation proposed by S. Haykin in [11]:

$$x_{i+m\tau} = F\{x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+m\tau-1}\}$$  \hspace{1cm} (12)

### 3. Network model principle

Recurrent neural network (RNN) is a kind of neural network with short-term memory capability, which is suitable for dealing with time-related problems. Long short-term memory network (LSTM) is a kind of RNN, which can solve the problem of long-distance dependence and gradient disappearance of RNN. LSTM and its variants are used to solve various problems, especially in text generation, speech recognition, handwriting recognition and so on.

Compared with one state normal RNN, there are four states in LSTM. The LSTM’s loop structure keeps a persistent unit state, which is used to decide whether the information should be forgotten or passed on. The internal diagram of LSTM’s loop structure is shown in Figure 1. Each loop structure has two outputs, where $C_i$ is the unit state and $h_i$ is the output of the network.

![Figure 1. The LSTM architecture.](image)

LSTM cells are composed of input gate, forget gate, output gate and cell state. The state update formula of each gate is as follows:

**Input gate:**

$$i_t = \sigma(x_tW_{xi} + h_{t-1}W_{hi} + b_i)$$  \hspace{1cm} (13)

**Forget gate:**

$$f_t = \sigma(x_tW_{xf} + h_{t-1}W_{hf} + b_f)$$  \hspace{1cm} (14)

**Cell state:**

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma(x_tW_{xc} + h_{t-1}W_{hc} + b_c)$$  \hspace{1cm} (15)

**Output gate:**

$$o_t = \sigma(x_tW_{xo} + h_{t-1}W_{ho} + b_o)$$  \hspace{1cm} (16)

**Final output:**

$$h_t = o_t \circ \sigma(c_t)$$  \hspace{1cm} (17)

Where $i_t$, $f_t$ and $o_t$ are the outputs of input gate, forget gate and output gate respectively; $c_t$ and $h_t$ are cell state and output respectively; $W$ and $B$ are weight matrix and bias matrix respectively.

Stacked LSTM uses multiple layers LSTM for stacking, in which the output of the upper layer is used as the input of the next layer.
Nested LSTM (NLSTM) is a LSTM architecture jointly proposed by CMU and the University of Montreal in 2017. In NLSTM, the memory cell of LSTM can access the internal memory and selectively read and write with standard LSTM gate. Instead of computing the value of the (outer) memory cell as equation (15), NLSTM memory cells use the memory function \( c_t = m_t (f_t \circ c_{t-1} \circ i_t \circ g_t) \). For the memory function, it can be realized by another LSTM cell. Similarly, the memory function can be replaced by another nested LSTM cell, so the nested network of complicated structure can be constructed.

The computational graphs of the three network models are shown in Figure 2.

![Figure 2. Computational graphs of the LSTM, Stacked LSTM, NLSTM.](image)

The sea clutter prediction model based on neural network is shown in Figure 3.

![Figure 3. Flow chart of sea clutter prediction.](image)

### 4. Experiment and result analysis

#### 4.1. Data introduction

The data used in this paper includes two parts: The first part of the data comes from IPIX radar data from McMaster University Laboratory in Canada in 1993. This experiment uses four sets of data #26, #54, #320, and #310 corresponding to 1-4 levels of sea state. Each group of data contains 14 range gates, including one main target gate, two or three secondary target gates, and the rest are sea clutter gates. Each range gate has 131,072 sampling points. The second part of the data is the sea clutter data collected by the P-band radar on the Yellow Coast of China. This experiment uses four sets of P-band data, which are P101, P201, P301, and P401 corresponding to 1-4 levels of sea state. Each group of data contains 100 range gates, all of which are sea clutter gates. Each range gate has 61,001 sampling points. In this experiment, the amplitude of sea clutter is predicted. So it is necessary to model and normalize the original data. The amplitude of sea clutter data of the fourth range gate in #26 and the first range gate in P101 are shown in Figure 4.
4.2. Experiments

The experiment uses the mutual information method and the Cao method to obtain the delay time $\tau$ and embedding dimension $m$ of each group of data, constructs the sea clutter prediction equation according to formula (12), and then uses the four network models of LSTM, Stacked LSTM, NLSTM and RBF to predict sea clutter and compare the prediction effects of the four networks. The experimental training set is 5000 sampling points, the validation set is 1500 sampling points, and the test set is 1500 sampling points. Choose the root mean square error function (RMSE) as the evaluation function to evaluate the prediction results:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (predicted_i - real_i)^2}$$

The first part of the experiment is to predict the sea clutter data of different range gates in the same file. The radar pulses used in different range gates are the same, and the sea conditions such as radar azimuth, wind speed and wind direction are similar. The experiment uses the #26 data of the IPIX radar data. The seventh gate of the data is the main target gate, the sixth and eighth gates are secondary target gates, and the rest are sea clutter gates. The prediction results of the network are shown in Table 1, where the data is the RMSE value predicted by the network, and $n$ is the number of hidden layers of the network. From Table 1, we can see that the network has relatively low prediction errors for the sea clutter data of the same file at different ranges, and this can meet the accuracy requirement of sea clutter prediction. The experimental results show that the network model has good prediction effect for all data of the same file.

The second part of the experiment predicts the sea clutter data corresponding to the four sea states. We use the #26, #54, #320 and #310 data of the IPIX radar data corresponding to 1-4 levels of sea state and the P101, P201, P301 and P401 data of the P-band radar data corresponding to 1-4 levels of sea state respectively. The experimental results are shown in Table 2.

From the experimental results in Table 2, it can be seen that when the sea state is high, the prediction error of IPIX radar sea clutter increases significantly, and the prediction performance of both the RBF and the three networks proposed in this paper has decreased. This is because in high sea state, the sea surface fluctuates violently and the internal mechanism of sea clutter is more complicated due to the influence of wind speed and other conditions, which leads to decline of the prediction effect of the network. The RMSE of the P-band radar sea clutter is lower than that of the IPIX radar sea clutter, and the prediction is more accurate. This is because the amplitude of the P-band sea clutter changes slowly.


### Table 1. Prediction error of each sea clutter gate in #26

| Model          | RBF  | LSTM | Stacked LSTM | NLSTM | Stacked LSTM | NSLTM |
|----------------|------|------|--------------|-------|--------------|-------|
| n              | 1    | 1    | 2            | 2     | 3            | 3     |
| 1              | 0.0307 | 0.0276 | **0.0248**  | **0.0248** | 0.0265 | 0.0265 |
| 2              | 0.0185 | 0.0157 | **0.0156**  | **0.0236** | 0.0262 | 0.0241 |
| 3              | 0.0318 | 0.0244 | 0.0246 | **0.0237** | 0.0276 | 0.0247 |
| 4              | 0.0257 | 0.0285 | 0.0255 | **0.0227** | 0.0235 | 0.0230 |
| 5              | 0.0269 | 0.0288 | 0.0236 | **0.0290** | 0.0344 | 0.0326 |
| 9              | 0.0545 | 0.0315 | 0.0309 | **0.0190** | 0.0211 | 0.0198 |
| 10             | 0.0191 | 0.0194 | 0.0180 | **0.0175** | 0.0182 | 0.0180 |
| 11             | 0.0396 | 0.0270 | 0.0272 | **0.0263** | 0.0279 | 0.0281 |
| 12             | 0.0322 | 0.0184 | 0.0204 | **0.0190** | 0.0211 | 0.0198 |
| 13             | 0.0315 | 0.0229 | 0.0238 | **0.0225** | 0.0248 | 0.0247 |
| 14             | 0.0150 | **0.0115** | 0.0118 | 0.0118 | 0.0115 | 0.0120 |

### Table 2. Sea clutter prediction results of four sea state

| Model          | RBF  | LSTM | Stacked LSTM | NLSTM | Stacked LSTM | NSLTM |
|----------------|------|------|--------------|-------|--------------|-------|
| n              | 1    | 1    | 2            | 2     | 3            | 3     |
| #26            | 0.0257 | 0.0285 | 0.0255 | **0.0237** | 0.0276 | 0.0247 |
| #54            | 0.0186 | 0.0167 | 0.0179 | **0.0167** | 0.0177 | 0.0169 |
| #320           | 0.0860 | 0.0460 | 0.0510 | **0.0447** | 0.056  | 0.0528 |
| #310           | 0.0085 | 0.0487 | **0.0485** | 0.0501 | 0.0539 | 0.0535 |
| P101           | 0.0893 | 0.0098 | 0.0097 | **0.0084** | 0.0168 | 0.0099 |
| P201           | 0.0122 | 0.0106 | 0.0129 | **0.0090** | 0.0143 | 0.0097 |
| P301           | 0.0044 | 0.0047 | 0.0091 | **0.0037** | 0.0095 | 0.0038 |
| P401           | 0.0114 | 0.0059 | 0.0060 | **0.0044** | 0.0070 | 0.0048 |

#### Figure 5. Prediction results of partial sea clutter data of IPIX radar and P-band radar

4.3. **Prediction performance analysis of network**

Based on the results of the above experiments, we compare and analyze the prediction performance of the network, and look into the factors that affect the prediction accuracy of the network.

1. **Comparison of prediction results of LSTM, Stacked LSTM, NLSTM and RBF**: Overall, the predicted RMSE of the three networks used in this paper are lower than those predicted by RBF, and
the prediction result is better than RBF. Especially in the prediction of sea clutter data with high sea states, the prediction performance of the three network is much better than that of the RBF. The experimental results show that the prediction performance of RNN (LSTM, Stacked LSTM, NLSTM) model is better than RBF in the prediction of sea clutter.

2. Comparison of the prediction effects of the three network models of LSTM, Stacked LSTM, and NLSTM:

1). For different hidden layers: From the prediction performance of one-layer LSTM, two-layer Stacked LSTM and three-layer Stacked LSTM model, the prediction performance of the network gradually decreases as the number of layers increases. Generally, the performance of one-layer LSTM is the most well, and the three-layer Stacked LSTM is the worst. For the NLSTM network, the predicted RMSE value of the two-layer NLSTM network is lower than the predicted RMSE value of the three-layer NLSTM. This result shows that for the prediction of sea clutter data, the increase in the number of network layers cannot bring about the improvement of the regression prediction effect for the prediction of sea clutter data. The increase in the number of layers will not only bring about exponential growth of time and memory costs, but also easily cause gradients to disappear. In addition, the model is a time series model, resulting in slowed update iterations of the LSTM layer close to the input layer and reduced convergence effects.

2). For the same hidden layer, the prediction effect comparison of Stacked LSTM and NLSTM: From the experimental results in Table 1 and Table 2, whether it is a two-layer NLSTM or a three-layer NLSTM, the predicted RMSE of the network is lower than the RMSE predicted by stacked LSTM with the same number of hidden layers. NLSTM increases depth by nesting instead of stacking, and can achieve a more efficient temporal hierarchy than traditional Stacked LSTM. Therefore, compared with Stacked LSTM with the same number of parameters, the prediction performance of NLSTM is better. Figure 6 shows the changes of RMSE values during training and verification of the three networks on the data of the fourth range gate in #26. It can be seen from Figure 6, all the RMSE values of the networks are decrease, and basically reach a relatively stable level after 50 times as the number of iterations increases. The RMSE convergence speed of the two Stacked LSTM networks is slower than the convergence speed of the others and the final RMSE value is larger. The RMSE of the two NLSTM networks converges fastest, and can converge to a lower value after less iterations. This also corresponds to Table 1 and Table 2. NLSTM has the lowest overall RMSE value and the best prediction performance on all test data sets.

Figure 6. Loss curve in the process of network training and verification

5. Conclusion

In this paper, we use phase space reconstruction to study sea clutter from the perspective of chaos and propose to use LSTM, Stacked LSTM and NLSTM to predict the amplitude of sea clutter. The experiment analyzes the prediction performance of the three networks from many aspects. And the
results show that the three networks used in this paper have higher prediction accuracy of sea clutter amplitude than traditional RBF. Since NLSTM is more suitable for processing long-term sequences in structure, the prediction effect is better than that of the other two.

The experimental results in this paper can also serve as reference for practical applications such as sea surface target detection. In the future, the network model will be improved in view of the problems existing in the network model, such as long training time, large prediction error and poor effect in high sea clutter prediction.

Acknowledgments
This work was supported by Equipment Pre-research Key Laboratory Fund Project(6142403190304).

References
[1] Haykin S and Puthusserypady S 1997 Chaotic dynamics of sea clutter Chaos An Interdiplinay Journal of Nonlinear ence 7(4): 777-802.
[2] Haykin S and Puthusserypady S 1997 Chaos, sea clutter, and neural networks Proc. of The 31st Asilomar Conf. on Signals, Systems and Computers p 1224-1227.
[3] Leung H and Lo P 1993 Chaotic radar signal processing over the sea IEEE Journal of Oceanic Engineering 18(3): 287-295.
[4] Xie N, Leung H, Chan H 2003 A multiple-model prediction approach for sea clutter modeling IEEE Trans. on Geoscience and Remote Sensing 41(6): 1491-1502.
[5] Ning Lu, Cai Yu and Wei Cai 2012 GRNN small target detection based on sea clutter Fire Control Radar Technology 41(2): 4-7.
[6] Ting Xu 2015 Prediction of sea clutter based on GA-WNN Electronic Design Engineering 23(18): 34-37.
[7] Jun Zhao, Jiaji Wu, Xiaowei Guo, Jie Han, Kai Yang, Hongguang Wang 2019 Prediction of radar sea clutter based on LSTM Journal of Ambient Intelligence and Humanized Computing DOI: 10.1007/s12652-019-01438-4
[8] Liwen Ma, Jinpeng Zhang, Jiaji Wu, Yushi Zhang, Peng Zhao and Xiaoyuan Xia 2020 Prediction of sea clutter using gated feedback recurrent neural network Chinese Journal of Radio Science 35(2): 257-263.
[9] Cers F A, Schmidhuber J, Cummins F 2000 Learning to Forget: Continual Prediction with LSTM Neural Computation 12(10): 2451-2471.
[10] Moniz J R A and Krueger D 2017 Nested LSTMs Proc. of Asian Conf. on Machine Learning (80): 1-15.
[11] Haykin S and Li X B 1995 Detection of signals in chaos Proc. of the IEEE 83(1): 95-122.