Solar Radio Filtering Algorithm Based on Improved Long Short-Term Memory

Qing-Fu Du, Qiao-Man Zhang, Xin Li and Chang-Lin Gao

School of Mechanical, Electrical & Information Engineering, Shandong University, Shandong, Weihai 264209, China; 201836556@mail.sdu.edu.cn.

Received 2020 June 25; accepted 2020 September 21

Abstract The effective observation of burst events in solar radio research has been impeded by various interference signals, especially interference signals with a wide frequency range and high intensity, as they can partially or completely obscure the observation of burst events. Image processing methods that directly remove the interference signal channels and subtract the average of the interference signal channel are not suitable for processing all types of interference signals. This paper proposes the use of a specific kind of recurrent neural networks, called long short-term memory networks, to predict the value of the RFI (radio frequency interference) signals with high intensity of the burst event in the solar radio spectrum. The predicted interference can then be removed in accordance with the principle that signals can be linearly added. Therefore, predicted value is subtracted from the data containing the burst event signals and the RFI signals (The radio frequency interference signals to be processed in this article refers to the signal of the broadcast signal that can be received in the frequency range, the signal transmitted by the mobile phone, and the signal transmitted by the sea vessel, and the like.) to remove the interference. Then, in order to reduce the error caused by the stepwise prediction in the network and further improve the prediction accuracy, this paper analyzes the characteristics of the value of the radio interference and applies the digital mapping method to convert the prediction problem into the classification problem in the time series. The experimental results show that the proposed method can effectively remove the radio interference in the solar spectrum and clearly show the burst events.

Key words: Sun: radio radiation — methods: data analysis — techniques: image processing

1 INTRODUCTION

Solar activity is closely related to our daily lives, so solar radio has become an important area in astrophysics research. In particular, the study of the solar radio burst process, carrying important information, not only helps to explain the physical process of the relevant plasma changes, but can also be used to find the law of energy change and analyze important physical phenomena such as material motion (Cheng 2018). In order to better research the detailed structure of solar radio bursts, the project team established a solar radio dynamic spectrometer with high time resolution and high frequency resolution at the Chashan Solar Observatory.

The Chashan Solar Observatory faces the sea and is far away from the town, which avoids a complicated communication environment and greatly reduces the radio frequency interference. However, there
are still some RFI signals and other electromagnetic interference in the meter band. These interference signals seriously affect the observation and analysis of solar radio burst events. In particular, when the interfering signals frequency is wide (the radio interference with a bandwidth greater than or equal to 2 MHz), and the signal intensity is greater than the intensity of the solar radiation flow, the interference can directly cover the burst event, so it is necessary to take anti-interference measures.

Common filtering countermeasures for solar radio spectrograms are divided into hardware processing and software processing methods. In hardware processing methods, Xu et al. use the spatial selectivity of the antenna to attenuate various interferences from the ground, use shielding, grounding, and amplifiers whose interference performance is good to suppress interference introduced by cable power, and use filters to suppress out-of-band signals, etc (Xu et al. 1995). Although the hardware circuit filtering method can filter out the interference signal, it cannot distinguish between a burst signal and an interference signal, which causes the loss of effective information. Therefore, the software processing methods are often used for image enhancement and interference removal. Most of these methods are used to process the gray-scale image of the spectrum. Common image processing methods for removing radio frequency interference is deleting the interference signal channels or subtracting the mean of its channels. The FFT or wavelet analysis can be used to remove the radio frequency interference, but this method is used to process the gray-scale image of the spectrum in most cases. Although they can enhance the image information, these methods will lose the details of a large number of burst events. In this research, we make use of cutting-edge technical means and a large of solar radio data, a deep learning method is selected to filter out the interference signals.

In recent years, deep learning methods such as multi-modal networks, long short-term memory networks (LSTM), deep-confidence networks, and convolutional neural networks all have good performance in the classification and archiving of solar radio spectrum and identifying burst events (Chen et al. 2015; Yu et al. 2017; Chen et al. 2016; Chen et al. 2017). For example, in terms of astronomical prediction, Huang X et al. use convolutional neural networks to predict solar flares (Huang et al. 2018); Sun W et al. use the LSTM model to predict the total electron content of the ionosphere through the temporal relationships within the data (Sun et al. 2017); Bikowski M et al. propose a significance-offset convolutional neural network (SOCNN) model for predicting multivariate asynchronous time series in combination with autoregressive models and cyclic neural networks (Bikowski et al. 2018). According to the research results available in the literature, we find that the interference signals to be removed in this paper is similar to the voice signals, and they have certain regularity in time (the time series characteristics mean that the signal is in a certain frequency range, and it shows a certain regular pattern in the process of random variation with time. When this discipline is found, the signal at the next moment can be predicted). Therefore, an algorithm for predicting the RFI signals value in the solar radio burst region that uses a recurrent neural network (RNN) is proposed. Then, according to the numerical characteristic (It means that the interference signals of each frequency channel are regularly changed in a fixed number of values, as exemplified in Section 2.3) of the radio interference, the algorithm is improved and digital mapping processing is added to improve the accuracy of the prediction.

2 RELATED THEORETICAL KNOWLEDGE

2.1 The Principle of Recurrent Neural Networks

An RNN accumulates data on the time axis using recursion to memorize information. Figure 1 is the structural diagram of the RNN, where \( x_t \) represents the input state at time \( t \), \( A \) represents the model processing part of the neural network (it is a repeating module with only one tanh layer), \( h_t \) is the state of the hidden layer at time \( t \), and \( y_t \) represents the output state at the corresponding time. The matrices of weight coefficients between the layers are represented by \( w_1, w_2 \) and \( w_3 \). Unlike the traditional neural network, the parameters \( w_1, w_2 \) and \( w_3 \) for the same location of the RNN at different times are shared, which greatly reduces the number of parameters for the required training. This reduction in parameters lowers the computational load and improves the network training speed.
As can be seen from the above figure, the hidden layer unit $h_t$ at the current moment receives the information of the hidden layer $h_{t-1}$ from the previous moment; the input layer information $x_t$ at the current moment passes data forward. Therefore, the output layer at each moment contains features of the past. The propagation process of the RNN can be expressed as:

$$h_t = f(w_1 x_t + w_2 h_{t-1} + b)$$  \hspace{1cm} (1)

$$y_t = f(w_3 h_t + b)$$  \hspace{1cm} (2)

where $h_{t-1}$ is the state of the hidden layer at the previous moment; $b$ is the bias term; and $f$ is the relationship of the nonlinear mapping, generally referred to as the activation function.

In theory, the RNN can process a sufficiently long time series. However, the activation function in the RNN often uses the tanh function and the sigmoid function, which leads to the phenomenon that the gradient disappears during the reverse transmission of the error (Yu et al. 2019; Zhang et al. 2017). The disappearing gradient makes the RNN only suitable for processing signals whose time series are short (Lv et al. 2015). In order to fully discover the time series characteristics and obtain more accurate predictions, the signal needs to be fully processed with a long time series (the time series signal for learning and training is approximately 30s and has at least 4600 frames of data, and the predicted time series signal is approximately 1.5s and has 150-200 frames of data. One frame of data represents the value of the interference signals for each channel at one point in time. For example, Fig. 3 is a solar radio frequency spectrum diagram drawn by a two-dimensional array data. Each row in the array represents data of a single frequency channel as a function of time, and each column represents data of all frequency channels at one point in time. Therefore, one frame of data is equivalent to one column of data in a two-dimensional array. Since the interference signals of a single frequency channel is processed here, one frame refers to the value of a single frequency channel at one moment.). Comparing two special RNN models: both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) effectively alleviate the problem of gradient disappearance. GRU converges quickly due to fewer parameters. However, in the case of a large data set, LSTM expresses better performance (Chung et al. 2014). Based on the current situation that the amount of data in this paper is sufficient, LSTM algorithm was finally selected.

### 2.2 The Principle of the Long Short-Term Memory

Different from the traditional RNN model, the LSTM model replaces the original hidden layer unit with the LSTM cell structural unit. This unit is illustrated in Figure 2 with grey, round edge rectangles. The information gate is generated by the sigmoid function and is added to the repeatedly linked module. It
consists of forgotten gates, input gates, and output gates (Klaus et al. 2016). The function property makes the information passing through each gate carry a parameter that controls the amount of information passed to the current neuron and the amount of information assigned to the next neuron.

\[
\begin{align*}
    f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
    i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
    o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
    c_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
    C_t &= f_t \cdot C_{t-1} + i_t \cdot c_t \\
    h_t &= o_t \cdot \tanh(C_t)
\end{align*}
\]

where \(f, i, o\) represent the information of the forgotten gates, the input gates, and the output gates, respectively. \(W\) refers to the weight coefficient matrix of the network, \(b\) is the bias term, \(x\) represents the input of the network, \(t\) represents the moment, \(h\) is the hidden layer state, \(c\) is the state value when the network updates the cell (it is also called the current candidate memory state value), and \(C\) is the current memory cell state value of the LSTM network (Li et al. 2018). The three information gates all use the sigmoid function as the activation function, and the tanh function is selected when the memory unit of the network is updated (Miao et al. 2016). The output information of the final hidden layer is related to the output value of the output gates and the current memory cell state value. In addition, compared with RNN, the backpropagation algorithm of the LSTM network not only calculates the error gradient corresponding to the hidden layer state \(h_t\), but also calculates the error gradient of the cell state \(C_t\) (Gers et al. 1999).

The application of three information gates in the LSTM network makes the memory cell in the network structure store historical information for a long time. From equation (7), it can be seen that the LSTM structural unit is composed of two parts, so when the cumulative error is calculated, the
result of zero is not present, which alleviates the problem of the gradient disappearance and realizing the long-term memory function.

The Gated Recursive Unit (GRU) is the latest development of the LSTM unit, both variants of the RNN. The GRU reduce the gating signals to two from the LSTM model. They are called an update gate and a reset gate (Rahul et al. 2017). Although the network structure of GRU is simpler than LSTM, experimental results show that LSTM performance is better than GRU in some cases of sequential prediction (Tolga et al. 2018), so this paper still chooses LSTM network structure.

3 METHOD AND APPLICATION

3.1 Data selection

A large solar burst event was observed on September 9, 2017, and the background value of the event in the quiet solar state is zero, so the data of that day are selected for de-interference processing. The data come from the high-frequency solar radio receiver of the Chashan Solar Observatory, with a frequency range of 150 MHz-500 MHz, a frequency resolution of 16 kHz, and a time resolution of 10 ms. Because the data received by the data acquisition card are a left-hand circular polarization signal and a right-handed circular polarization signal after the digital polarization synthesis operation, the two signals exist in two data channels. Their values are similar, so only one of the signals, the left-handed circular polarization signal, is selected in this paper for drawing and processing images.

\[
\text{Thresh} \leq \frac{(I(t))^2}{\text{Var}(I(t))}
\]  

where \( I(t) \) represents the signal strength value over time in a single channel, \( (I(t))^2 \) and \( \text{Var}(I(t)) \) represent the mean and standard deviation of all signal strength values within the \( t \) period, respectively.
The specific size of the threshold for screening the RFI signals, Thresh, can be selected according to the actual situation, and the value of thresh in this article is 40.

In order to more accurately divide the training set in the neural network, it is also necessary to locate the time when the burst event starts. This paper takes the method of averaging the intensity values of multiple channels of the target RFI signals and transforming it into a one-dimensional array that changes with time. Take the radio interference in the 360 MHz-380 MHz band as an example. As shown in Figure 4, when the sun has not erupted, the average value of each channel is the average value of the RFI signal which experiences little change. However, when the burst events occur, the average value is the average value of the radio interference and the burst value, which is significantly higher than the pure radio interference average. The start time of the burst event handled in this article is the position pointed by the arrow in Figure 4.

3.2 Prediction of Radio Interference Signal Value Based on LSTM Model

3.2.1 Building a data set

The establishment of data sets and network training are the two main parts of deep learning. The purpose of data set standardization is to eliminate the dimensional influence between data, transform data with different units or magnitudes into non-dimensional values, and realize analysis and comparison under the same magnitude. Commonly used methods include normalization. After data standardization, a data set must be established. In order to reduce the amount of network training tasks and improve the accuracy of prediction, the sequence is segmented, and a mapping is considered to be established to find the relationship between the time period and the time period. The process of establishing the data set is shown in Figure 5.

The picture is divided into three parts: input window, output window, and sliding window. A sliding window contains a complete input window and output window. The three windows as a whole slide in the time series, and the sliding unit is one. The input window and the output window contain the same amount of data which is named the time step; this is used as one of the network hyperparameters. the length of the sliding window is related to time step.once determine the time step is determined and also the length of sliding window is determined.In addition, in order to preserve the connection between adjacent sequences as much as possible and improve the accuracy of prediction, the output window and the input window have only one time series interval. Therefore, the sequence values (t-time step + 2) to (t + 1) are predicted from the (t-time step + 1) to (t) sequence information.
3.2.2 Establishment of LSTM network and prediction of radio interference values

The LSTM network built in this paper is divided into five layers (Figure 6): input layer, fully connected layer 1, LSTM layer, fully connected layer 2, and output layer. Firstly, the data set is transformed into the required dimension of the network after entering the first fully connected layer. Then, the LSTM layer automatically extracts a series of timing features through its internal loop operation. Then, the output feature vector enters the second full connection layer. Finally, the network outputs the predicted radio interference value. As can be seen from Section 3.2.1, the last frame of data output by the network is the radio value for the next moment of prediction.

In order to improve the network performance and enhance the learning ability of the model, it is generally preferred to increase the number of network layers and the number of nodes in the layers. However, increasing the depth of the model not only increases the time of the network training, but also causes over-fitting that reduces the final prediction accuracy. Therefore, this section of the experiment selects four LSTM layers to extract features, and the number of nodes is 25 (The parameters are based on the experimental results in Table 1). In addition, the Back Propagation Through Time (BPTT, the principle of BPTT algorithm is consistent with Back Propagation, BP. The difference is that when calculating the
error term, BPTT is necessary to calculate the error in a time direction to update the weight between hidden layers. The BP algorithm consists of two processes: forward propagation of the signal and back propagation of the error.) algorithm and the gradient descent algorithm are selected to complete the training of network weights. An MSE loss function and a mini-batch based stochastic gradient descent optimizer are added according to the training results.

After the model training is completed, it enters the prediction stage. In order to test the performance of the network, the test data (the value of the known pure radio interference signal position) is used for prediction. The network structure is adjusted by comparing the results using the original data of the corresponding position. Since the position of the radio interferences value to be predicted is unknown, a step-by-step prediction method is required for prediction, i.e., the predicted value is added to the next input for a second prediction every time. Figure 7 shows a comparison of the 100-frame radio interference signal predicted by a single channel with the original data. The blue line indicates the test set data and the red line indicates the prediction result. It can be seen from the figure that the discipline of the predicted value and the actual value change with time is about the same, but the numerical values are different. The reason is that step-by-step prediction method introduces errors and the prediction ability of the model decreases over time.

![The prediction result of the test set](image)

**Fig. 7: The prediction result of the test data (100-frame radio interference signal)**

### 3.3 Prediction of radio interference value based on improved LSTM network

As can be seen from the previous section, the step-by-step prediction method continuously introduces errors, which affects the overall prediction results. By observing the characteristics of the radio interference signal, we find that the data of each channel has only a fixed number of values. Therefore, adopting the method of adding a digital mapping is proposed to reduce the error and improve the accuracy of prediction. For example, Figure 8 shows the original sequence diagram and the histogram of 6000 frames of data from four randomly selected channel. Figure 8 shows that the values of these channels have a different number of categories, ranging from 2 to 5. Therefore, the values appearing in the sequence can be treated as different classes, and the prediction problem is converted into a classification problem. The predicted value is the value with the highest predicted probability in the classification. As long as the predicted trend is correct, the accuracy of the predicted value at that moment is 100%. This avoids the problems illustrated in Figure 7 (We processed the data entering the network based on the Z-score normalization method. Therefore, the data shown in Fig. 7 is different from that in Fig. 8).
3.3.1 Building a data set

The establishment of the dataset in this section is basically the same as that in Section 2.2. Data preprocessing and a sliding window are required to extract the data. The difference is that the size and position of the output window in the sliding window has changed, and the data in the window is labeled. The schematic diagram of the changed data set is shown in Figure 9. The output window contains only one frame of data and is located next to the end of the input window. Therefore, the sequence values of \((t + 1)\) are predicted from \((t\)-time step + 1) to \((t)\) sequence information, while the sliding window still slides only one frame at a time. After obtaining the value by segmentation sliding, the mapping of the time segment (input window) to the time point (output window) is generated. Then, the classification corresponding to the time point data is converted into the label of the corresponding time segment.

The data in the sliding window are labeled by One-Hot encoding which converts the number of categories into binary numbers. The N-type values are encoded with an N-bit register to ensure that only one valid position per class is activated and marked as one; the other locations are marked as zero. For example, the values of the radio interference signal in Figure 8(b) have four categories, so \(N = 4\).

3.3.2 Prediction of the Radio Interference Value Based on the Digital Mapping Method

The network structure used in this section is similar to Section 3.2. However, for better classification the dropout function is added to the model to avoid over-fitting in the network, and the loss function is
Table 1: Experimental results of parameter adjustments - 3000 frames

<table>
<thead>
<tr>
<th>num_layer</th>
<th>unit</th>
<th>accuracy</th>
<th>the accuracy after adding dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>0.731</td>
<td>0.735</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>0.728</td>
<td>0.731</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.736</td>
<td>0.74</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>0.72</td>
<td>0.735</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0.746</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>0.740</td>
<td>0.734</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.738</td>
<td>0.744</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>0.73</td>
<td>0.747</td>
</tr>
</tbody>
</table>

changed to the softmax cross-entropy function (Hu et al. 2018; P et al. 2016;), which is more suitable for classification operations. The learning and updating of weights utilizes the Adam algorithm (Chang et al. 2019) for the adaptive learning rate optimization.

Considering that this paper needs to process 768 channels of RFI signals, it means we should call 768 network models for prediction. The hyperparameters involved in this paper have time step length, the number of hidden layer node unit, the number of trainings entering the network each time batch size, network depth num_layer, and learning rate lr, etc. The workload is heavy when each channel adjusts the hyperparameters to obtain the optimal model, since the training process for a single channel takes approximately 30 minutes, this means that each modification of the hyperparameter must be retrained for 30 minutes, while the 5 hyperparameters of the 768 channel require at least 5 * 768 * 0.5 = 1920h. Therefore, a large number of experiments are first used to screen out a set of optimal hyperparameters for the learning, training, and prediction of the 768 models.

The sample set for screening the optimal hyperparameters consists of 10 channels of data that are randomly selected from the 768 RFI signal channels. In order to determine the appropriate training set, experiments with samples of 3000, 6000 and 9000 frames were carried out, and the data was divided into a test set (5%) and a training set (95%). The idea of dichotomy is used here. First try to use the smaller number of frames 3000 and the larger number of frames 9000, because the data of the observation system is about 10000 frames per unit stored in the disk array. Through experiments, it is found that too small amount of data and too much data amount are not conducive to network training. The experimental results are as: Table 1 and Table 2.

From the data recorded in the table, it is found that the accuracy of 9000 frames of data is lower than that of 3000 frames of data. It can be estimated that the appropriate number of frames must be between 3000 and 9000, so it is better to determine 6000 as the appropriate number.
Table 2: Experimental results of parameter adjustments - 9000 frames

<table>
<thead>
<tr>
<th>num_layer</th>
<th>unit</th>
<th>accuracy before dropout</th>
<th>accuracy after adding dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>0.706</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>0.710</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.704</td>
<td>0.72</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>0.714</td>
<td>0.725</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0.726</td>
<td>0.733</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>0.730</td>
<td>0.72</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.733</td>
<td>0.735</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>0.738</td>
<td>0.732</td>
</tr>
</tbody>
</table>

Table 3: Experimental results of parameter adjustments - 6000 frames

<table>
<thead>
<tr>
<th>num_layer</th>
<th>unit</th>
<th>accuracy before dropout</th>
<th>accuracy after adding dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>0.796</td>
<td>0.800</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>0.768</td>
<td>0.804</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>0.800</td>
<td>0.801</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>0.776</td>
<td>0.808</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.784</td>
<td>0.808</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>0.704</td>
<td>0.800</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>0.764</td>
<td>0.778</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>0.796</td>
<td>0.796</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0.776</td>
<td>0.796</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>0.800</td>
<td>0.812</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>0.780</td>
<td>0.800</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.448</td>
<td>0.784</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>0.752</td>
<td>0.760</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>0.748</td>
<td>0.796</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>0.764</td>
<td>0.808</td>
</tr>
</tbody>
</table>

The experimental results of some parameter adjustments are presented in Table 3. The effects of network depth, the number of hidden layer nodes, and dropout function on network prediction are studied under the condition of a fixed learning rate (lr=0.001) and time step (time_step=35). The numerical values of the learning rate and time step are one of the results randomly selected from the exhaustive way of traversing all the parameter combinations.

As can be seen from the table, it is not the deeper, the better for the network. When the number of nodes is 20 and the network has five layers, the accuracy of the prediction is only 44.8%. This is because the network structure has been over-fitting with the deepening of the network. The prediction accuracy is improved to 78.4% after adding the dropout function. The prediction accuracy after adding the dropout function is consistently higher than that without adding the function, which demonstrates the value of adding the dropout function. From the table, a set of hyperparameters with the highest prediction accuracy rate is selected; these are the 4-layer network with 25 nodes. These values are used to continue the experiment for the learning rate and time step. According to the results of the table2, finally, the learning rate in the optimal hyperparameter is determined to be 0.001, and the time_step is 20.

4 THE RESULTS AND ANALYSIS OF THE INTERFERENCE SIGNAL SUPPRESSION

4.1 Prediction of radio interference value based on improved LSTM network

The actual value of the radio interference signal to be processed in this paper is unknown in the burst position. In order to verify the effect of using the method in this paper and the performance of the digital mapping method in the filtering process, we establish the event simulation. The two methods mentioned in Chapter 2 are used to process and analyze the event simulation.
Table 4: Experimental results of parameter adjustments: learning rate and time step (4 layers and 25 nodes)

<table>
<thead>
<tr>
<th>learning rate</th>
<th>time step</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.804</td>
</tr>
<tr>
<td>0.1</td>
<td>20</td>
<td>0.802</td>
</tr>
<tr>
<td>0.1</td>
<td>25</td>
<td>0.804</td>
</tr>
<tr>
<td>0.1</td>
<td>30</td>
<td>0.800</td>
</tr>
<tr>
<td>0.1</td>
<td>35</td>
<td>0.804</td>
</tr>
<tr>
<td>0.01</td>
<td>15</td>
<td>0.804</td>
</tr>
<tr>
<td>0.01</td>
<td>20</td>
<td>0.808</td>
</tr>
<tr>
<td>0.01</td>
<td>25</td>
<td>0.796</td>
</tr>
<tr>
<td>0.01</td>
<td>30</td>
<td>0.792</td>
</tr>
<tr>
<td>0.01</td>
<td>35</td>
<td>0.792</td>
</tr>
<tr>
<td>0.001</td>
<td>15</td>
<td>0.804</td>
</tr>
<tr>
<td>0.001</td>
<td>20</td>
<td>0.812</td>
</tr>
<tr>
<td>0.001</td>
<td>25</td>
<td>0.804</td>
</tr>
<tr>
<td>0.001</td>
<td>30</td>
<td>0.708</td>
</tr>
<tr>
<td>0.001</td>
<td>35</td>
<td>0.776</td>
</tr>
</tbody>
</table>

According to the principle that signals can be linearly added, the simulation data are superimposed with the pure burst data and the pure RFI data. The pure burst data are composed of the data from 6:52:23:461.7-6:52:27:131.7 (international time) and 225 MHz-240 MHz on September 9, 2019, which has 471 channels and 351 frames. The pure radio data consists of the data on September 9th, 2017 at 6:49:48.262-6:51:12.808.6 and 248.85 MHz-249 MHz, which has 6 channels and 8064 frames. Keeping the frequency band of the burst event unchanged and the time of the radio interference unaltered, the pure RFI signal value is superimposed to 227.5 MHz, and then it is doubled to the position of 231 MHz. Consequently, we obtain a simulation burst event with a frequency range of 224 MHz-247 MHz and a time range of 6:49:48.262-6:51:12.808.6. This is shown in Figure 10.

Firstly, the LSTM network is used to train the six selected channels of radio interference data and predict the RFI signal value of the burst position. Then, the predicted burst position data is subtracted to obtain the processing result graph shown in Figure 11. What can be seen from the processed intensity map is that some visible interference signals remain in the burst event, and the results are not very satisfactory.

The RFI signal is processed using the LSTM network based on the digital mapping method. Table 3 shows the accuracy of the prediction results of the six channels after classification and labeling. Although some channels accuracy is as low as 0.5 (The reason for the low accuracy is that the more...
the number of label classifications, the more complicated the signal variation law, so we need to train with more data to get higher accuracy. The average accuracy of the overall data can reach 74.33%. The final intensity map is shown in Figure 12. The RFI signals in the burst area are mostly removed, and the observation of the event is not obscured. Comparing the result with original image without RFI (Figure 11(c)) and Figure 11(b), the filtering effect of the LSTM network (Figure 12) based on the digital mapping method is significantly better than that of the general LSTM network.
4.2 Analysis of actual event processing results

4.2.1 Compensation for actual event processing

The simulation event superimposes the pure RFI signal value and the intensity value of the pure burst event, which conforms to the principle that signals can be linearly added. However, we found that the processing result is excessively suppressed. The actual RFI signal value of the burst position is smaller than the predicted value, and a part of the burst information is removed when the radio predicted value is subtracted from the original value of the burst position. Figure 13 is the result of the radio interference processing of a small burst event that occurs two minutes before the event in Figure 3. It can be seen that when the improved LSTM network is used to predict the radio interference value near 244 MHz and 253 MHz and process the event, there is a gap between the actual effect and the theoretical result. This result is due to a sharp increase in the signal strength when the solar radio bursts. Once the signal is amplified, the amplifier may work in a nonlinear region, resulting in the attenuation of the collected data. In addition, the process of the solar burst also affects the radio interference signal to some extent. Therefore, the two signals no longer satisfy the linear additivity of the signal, and the processing needs to be compensated in some way.

The system used to collect data is a high frequency resolution system with an average of 32 frequency channels corresponding to 1 MHz. The solar burst events usually occupy a large bandwidth in the frequency domain and are distributed in hundreds of frequency channels of the system. The high frequency resolution makes the values in adjacent frequency channels similar, which makes it possible to compensate for the processing results using burst data from adjacent radio interference signal channels. The compensation calculation formula is as follows:

\[ F(a,k) = \frac{f(a) + k \times g(b)}{2} \] (10)

where \( F(a,k) \) is the compensated value, \( f(a) \) is the value of the area where the interference signal is over-attenuated, \( g(b) \) is the value of the adjacent channel of the interference signal, \( k \) is the coefficient, and \( a \) and \( b \) represent different frequency channels. The coefficient \( k \) is selected according to the actual situation. Figure 14 (a), (b) are the compensated images of Figure 13 (c), (d), respectively. The compensated images better preserve the effective information of the burst event while removing the radio interference.

4.2.2 The processing result of actual event

The actual burst event whose time range is 6:54:4.617.8-6:54:46.854.4 is Figure 3 in Section 2.1, and the prediction time starts at 6:54:30.507.2. In order to further prove the superiority of the proposed
method, the ordinary image processing method is firstly adopted for the burst event to remove the radio interference signals. We choose the method of subtracting the mean of the single channel radio interference signals. To avoid the influence of the burst value on the channel mean, we select the 500 frames of data around the event 6:54:15 to calculate the channel mean, then the time and frequency resolution are reduced by six times and 16 times, respectively. The results are shown in Figure 15. It can be seen from the figure that the burst continuity of the wide red frame portion covered by the radio interference signals is destroyed, and the effective information is removed. Therefore, this section uses the improved LSTM network to process the radio interference signals in the outburst location; the interference signals that are not related to the burst event and the weaker interference signals are processed by subtracting the mean of the corresponding channels value.

Figure 16 is the processed result of using the method that combines the improved LSTM network and compensation, where k is taken as 20. In comparison with the original image and Figure 15, we can see that not only small-scale burst events between 250 MHz and 300 MHz can be clearly displayed in the spectrum, but also the burst event that is originally covered in the red frame can be clearly observed. Furthermore, the continuity of the burst events in frequency is also preserved.

In order to more intuitively compare the two processing methods and further demonstrate the superiority of using the methods herein to remove solar radio interference, some details of Figures 15 and 16 are shown simultaneously in Figures 17, 18, and 19. Part of the spectrogram allows for a clear view of the processing of each radio interference (multiple channels). Where Fig.a is the result of processing using the improved LSTM method, Fig.b is the result of the processing of the ordinary image processing method. The information in Fig.a is more reserved than the Fig.b about the burst signal.

In summary, the intensity of the interference signal at the burst is unknown and the accuracy of the network prediction cannot be determined. However, according to the experimental results, it can be found that regardless of the frequency band occupied by the interference signal, the proposed method can be used to process the radio interference signal. Then, the Binkowski2018Autoregressiveprocessed part is compensated by using the degree burst information. In this paper, the interference signal is finally removed while the burst information is retained as much as possible.

Fig. 13: Compensated images in 244 MHz and 253 MHz
5 CONCLUSION

In this paper, a novel method of predicting the radio interference signals in the solar radio spectrum based on the LSTM network is proposed. According to the characteristics of the RFI signals, the long time series signals are mapped to a certain time step to find the relationship between time periods. The predicted radio trend is similar to the original data, but the predictions are not accurate enough. In view of this situation, we propose an improved LSTM network based on the digital mapping method. Before entering into the network, the single frequency channel radio data should be classified. First, we establish the mapping from time periods to time points. Then, the corresponding time period label is established, which is according to the classification of the output time point. The improved LSTM network greatly improves the accuracy of the prediction results and provides a possibility of lossless filtering of the solar spectrum.

However, to remove the radio interference to achieve better results and to ensure a higher prediction accuracy, the use of this method has certain limitations, because the data in this paper is received by the high-resolution spectrum analyzer. There are fewer data classifications in each frequency channel, and the data of adjacent frequency channels does not change significantly. If the prediction of low-resolution data is performed, the accuracy of prediction will be reduced because the data classification is increased.

The method mentioned in this paper still has room for optimization: it takes substantial time to train the 768 networks in this event. A GPU can be used to train and predict multiple networks at the same time to improve work efficiency. In addition, although the errors caused by stepwise prediction are effectively suppressed by the methods in this paper, the error itself cannot be avoided. Finally, more periodic data can be added to support the analysis and extraction of additional valid features, and the use of step-by-step predictions can be avoided.


References

Alex, G., Navdeep, J., & Abdel-Rahman, M. 2013, in Automatic Speech Recognition & Understanding
Binkowski, M., Marti, G., & Donnat, P. 2017
Chen, S., Xu, L., Ma, L., Zhang, W., & Yan, Y. 2017, in IEEE International Conference on Multimedia & Expo Workshops
Chen, Z., Ma, L., Xu, L., Tan, C., & Yan, Y. 2016, Multimedia Tools & Applications, 75, 2859
Chen, Z., Ma, L., Xu, L., Weng, Y., & Yan, Y. 2015, in IEEE International Conference on Systems

![Fig. 14: Compensated images in 244 MHz and 253 MHz](image_url)

(a) Compensated image in 244 MHz  
(b) Compensated image in 253 MHz
Fig. 15: The intensity map of the ordinary image processing results

Fig. 16: The intensity map of the combination processing results (improved LSTM and compensation)

Cheng, R. 2018, The digital polarization solar radio spectrum with high resolution, PhD thesis, Shandong University
Li, Z., He, D., Tian, F., et al. 2018
Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. 2015, IEEE Transactions on Intelligent Transportation Systems, 16, 865
P. S., J. T., & K. B. 2016, Expert Systems with Applications, 47, 30463059
Rahul, D., & Fathi M., S. 2017
Yu, L., Qu, J., Gao, F., & Tian, Y. 2019, Shock and Vibration, 2019, 1
Yu, X., Xu, L., Ma, L., Chen, Z., & Yan, Y. 2017, in IEEE International Conference on Multimedia & Expo Workshops
CSO 2017−09−09 Improved LSTM processing method

CSO 2017−09−09 Ordinary image processing method

Fig. 17: 365MHz-385MHz interference radio processing results

Zhang, Y. 2017, Activation function oriented RNN algorithm optimization, PhD thesis, Zhejiang University
Fig. 18: 215MHz interference radio processing results
Fig. 19: 244MHz interference radio processing results