RAA 2020 Vol. **20** No. **1**, 7(8pp) doi: 10.1088/1674–4527/20/1/7

http://www.raa-journal.org *http://iopscience.iop.org/raa*

Research in Astronomy and Astrophysics

Real-time abnormal light curve detection based on a Gated Recurrent Unit network

Rui-Qing Yan[1], Wei Liu[1], Meng Zhu[1], Yi-Jing Wang[1], Cong Dai[1], Shuo Cao[2], Kang Wu[1], Yu-Chen Liang[1], Xian-Chuan Yu[1] and Meng-Fei Zhang[3]

[1] College of Information Science and Technology, Beijing Normal University, Beijing 100875, China; *yuxianchuan@163.com, 201822210015@mail.bnu.edu.cn*

[2] Department of Astronomy, Beijing Normal University, Beijing 100875, China; *caoshuo@bnu.edu.cn*

[3] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; *zmf@bao.ac.cn*

Received 2019 May 31; accepted 2019 July 29

Abstract Targeting the problem of high real-time requirements in astronomical data processing, this paper proposes a real-time early warning model for light curves based on a Gated Recurrent Unit (GRU) network. Using the memory function of the GRU network, a prediction model of the light curve is established, and the model is trained using the collected light curve data, so that the model can predict a star magnitude value for the next moment based on historical star magnitude data. In this paper,we calculate the difference between the model prediction value and the actual observation value and set a threshold. If the difference exceeds the set threshold, the observation value at the next moment is considered to be an abnormal value, and a warning is given. Astronomers can carry out further certification based on the early warning and in combination with other means of observation. The method proposed in this paper can be applied to real-time observations in time domain astronomy.

Key words: methods: data analysis — techniques: photometric — stars: variables: general

1 INTRODUCTION

In the development of astronomy, time domain astronomy has become an important development direction (Graham et al. 2012). It is based on high-resolution observations to anticipate, discover and study of extreme, rare astronomical phenomena in the universe that are related to cutting-edge astrophysics, such as extreme physical processes, dark energy, and the origin of life. In addition, these anomalous astronomical phenomena can also help scientists detect unusual, rare or unknown astronomical objects and phenomena (such as high redshift quasars, brown dwarfs, pulsars, etc.) (Dalcanton et al. 1994). However, this astronomical phenomenon of time domain variation is not easily detected with traditional observing equipment, and it requires a sky survey with high time domain resolution (Condon et al. 1998). GWAC (the Ground-based Wide-angle Camera array) is part of the SVOM (Space Variable Objects Monitor) of Sino-French cooperation. GWAC is exposed every 15 seconds and can get millions of light curves throughout the day. However, the analysis and processing of GWAC observation data faces many challenges, mainly including: (1) A large amount of data. For example, the amount of data reaches the level of TB, and it will grow rapidly in the later stages. At the same time, the real-time requirements are higher, that is, the processing efficiency of the data has higher requirements. (2) The data is diverse. Since the varying periods and brightness of different variable stars are different, the form of anomalies is different, making the data diverse. In the time domain signal processing tasks (Battistelli et al. 2008), one of the important tasks is the real-time early warning task of the variable sources. The scientific goal is to detect the abnormal light curves. Real-time warnings not only have a wide range of applications in the field of time domain astronomy, but also in other fields such as geological disasters (Zhang et al. 2005), financial crises in the economic field, and crisis events in the field of power grid security. At present, abnormal light curves are detected from massive datasets, which creates fatigue for astronomers, thus increasing the probability of misjudgment. However, GWAC puts higher demands on the efficiency and accuracy of light curve anomaly detection. Therefore, it is imperative to study faster and more effective abnormal light curve detection. How to detect abnormal celestial signals (such as transient sources) from the massive light curve and give real-time warning is an important technical problem in real-time monitoring. An example is the US LSST (Large Synoptic Survey Telescope), whose scientific goal is to search for optical transient sources (Sguera et al. 2006). At present, the main idea of the identification of the variable source is to continuously collect the data from the telescope for the same sky area within a specified period of time, and then to find the source of the change and the source of the transient according to the obtained data. The specific methods mainly include star table matching and image subtraction. The star table matching method mainly matches the observed star table with a template star table (Bhatti et al. 2010; Telezhinsky et al. 2010). If a star appears in the observed star table and there is no relevant record in the template star table, then this new emerging star is further studied to determine if it is the source of the transients being sought. Another method of detecting transient sources is where the image is poor and the candidate for the transient source is looked up based on the image residual map (Alard 2000; Bramich 2008). The methods of star table matching and image difference are simple and easy, but the two methods are relatively ineffective, and are particularly sensitive to noise in the observations. The most important thing is that they do not meet the real-time requirements of current time domain astronomy.

Recently, machine learning has made significant progress in the fields of computer vision (Lecun et al. 2015; Zhao & Du 2016; Zhang et al. 2015), natural language processing (Young et al. 2018; Sarikaya et al. 2014), speech recognition (Hinton et al. 2012; Abdel-Hamid et al. 2014), etc. Machine learning has also been widely used in astronomical data processing (Liu et al. 2019; Jones et al. 2017; Hon et al. 2017). However, recurrent neural networks (Zaremba et al. 2014) have shown strong advantages in many machine learning tasks, especially when the input or output has variable length characteristics. Although the regression problem is a hot topic in the field of computer research, there are few studies on the characteristics of data in astronomy and the regression prediction for specific scientific needs. In order to meet the real-time requirements for variable source detection, the main idea for a real-time early warning system is based on the prediction of time series signals. For example, using the prediction model of the Autoregressive Integrated Moving Average (ARIMA) (Dickey 1984), the light curve in a future period is predicted. If the actual observation exceeds a certain range of the predicted value, it is considered to be an abnormality. This method is sensitive to the length dependence of the window. Later, some scholars proposed a light curve anomaly detection algorithm based on the Recurrent

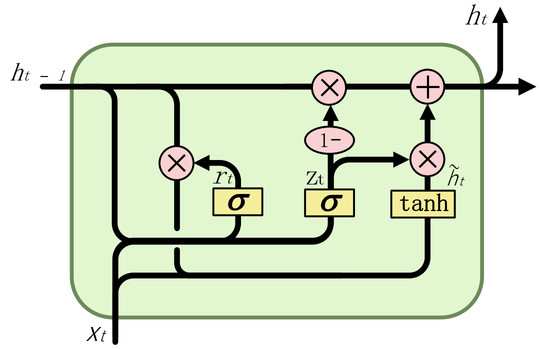


Fig. 1 GRU Structure. z_t and r_t represent the update gate and reset gate respectively.

Neural Network (RNN) model (Naul et al. 2018). The idea in this paper is similar to ARIMA. It also learns the data in a known period of time, and then predicts the data in the subsequent period based on the known data. However, RNN will produce a gradient disappearance during the training process, resulting in many parameters not being effectively returned.

Based on the above analysis, the most important point about real-time anomaly warning is establishing a prediction model that represents normal or abnormal samples. However, due to the limited amount of abnormal data, this paper starts from the establishment of a model that can effectively learn the characteristics of normal samples, and proposes an anomalous light curve early warning model based on a Gated Recurrent Unit (GRU). The model learns the characteristics of the normal light curve by means of neuron connections and predicts the trend of the normal light curve.

Our paper is structured as follows. The basis algorithm is described in Section 2. In Section 3, we describe data preprocessing and the structure and algorithm of the real-time early warning model. Section 4 discusses experimental results obtained from the model proposed from Section 3. Finally, conclusions are presented in Section 5.

2 BASIS ALGORITHM

GRU (Cho et al. 2014) is an abbreviation for Gated Recurrent Unit, which is in fact a variant of Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber 1997). The LSTM consists of three gates: the forget gate, the input gate, and the output gate. The GRU consists of only two gates, the update gate and the reset gate, as shown in Figure 1.

z_t and r_t in Figure 1 represent the update gate and reset gate of the GRU, respectively, where the update gate is used to control how much of the previous state information is involved in the current state. The larger the value of z_t, the more information from the previous state is included in the current state operation. The smaller the value of the reset gate r_t, the more the information of the previous state is ignored. From the GRU graph, we can find the relationship between the variables in the graph. The expression for the network forward propagation is as follows:

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]), \\ z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]), \\ \tilde{h_t} &= \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]), \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h_t}, \\ y_t &= \sigma(W_o \cdot h_t). \end{aligned} \tag{1}$$

The [ ] in the expression represents the vector connection, and $*$ represents the multiplication of the elements in the matrix. According to Equation (1), in the GRU training phase, the parameters that need to be optimized and learned in the backward propagation are W_r, W_z and $W_{\tilde{h}}$, i.e.:

$$\begin{aligned} W_r &= W_{rx} + W_{rh}, \\ W_z &= W_{zx} + W_{zh}, \\ W_{\tilde{h}} &= W_{\tilde{h}x} + W_{\tilde{h}h}. \end{aligned} \tag{2}$$

Suppose that the input of one of the output layers is $y_t^i = W_o h$, and the output is $y_t^o = \sigma(y_t^i)$, then the loss function at time t is $E_t = \frac{1}{2}(y_d - y_t^o)^2$. Then the overall loss function for a single sample is $E = \sum_{t=1}^{T} E_t$. According to the chain derivation rule, we optimize the parameters in the network, and the derivation process is as follows:

$$\begin{aligned} &\frac{\partial E}{\partial W_o} = \delta_{y,t} h_t, \quad \frac{\partial E}{\partial W_z x} = \delta_{z,t} x_t, \\ &\frac{\partial E}{\partial W_z h} = \delta_{z,t} h_{t-1}, \quad \frac{\partial E}{\partial W_{\tilde{h}x}} = \delta_t x_t, \\ &\frac{\partial E}{\partial W_{\tilde{h}h}} = \delta_t (r_t \cdot h_{t-1}), \\ &\frac{\partial E}{\partial W_{rx}} = \delta_{r,t} x_t, \quad \frac{\partial E}{\partial W_{rh}} = \delta_{r,t} h_{t-1}, \\ &\delta_{y,t} = (y_d - y_t^o) \cdot \sigma', \\ &\delta_{h,t} = \delta_{y,t} W_o + \delta_{z,t+1} W_{zh} + \delta_{t+1} W_{\tilde{h}h} \cdot r_{t+1} \\ &\qquad + \delta_{h,t+1} W_{rh} + \delta_{h,t+1} \cdot (1 - z_{t+1}), \\ &\delta_{z,t} = \delta_{t,h} \cdot (\tilde{h}_t - h_{t-1}) \cdot \sigma', \\ &\delta_t = \delta_{h,t} \cdot z_t \cdot \phi', \\ &\delta_{r,t} = h_{t-1} \cdot [(\delta_{h,t} \cdot z_t \cdot \phi') W_{\tilde{h}h}] \cdot \sigma'. \end{aligned} \tag{3}$$

3 PROCESS OF ABNORMAL WARNING ALGORITHM

The GRU model is a common regression prediction model. In the process of modeling the normal light curve, the GRU-based network model is robust to normal light curve prediction. However, we need to give real-time warning for an abnormal light curve in the actual problem, so the idea adopted in this paper is to use the GRU model to predict the light curve in the future. Then, according to the model, the threshold of the historical light curve learning result is given. The threshold is completely driven by data. If an observation from the telescope for a period of time in the future exceeds the set threshold, the signal is considered to be an anomaly signal and an early warning is given.

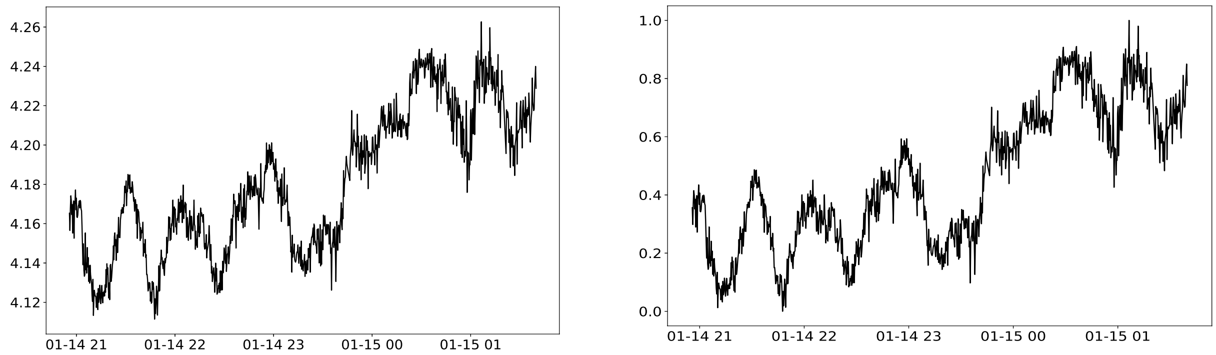
3.1 Data Preprocessing

The original light curve observation signal is easily polluted by light, so sometimes the overall trend of the signal we observe is rising or falling, that is, the range of observations in the normal light curve is relatively large. In order to speed up the model training, we first normalize the original observation data and scale the range of the original observations to [0,1]. Figure 2 shows the raw observation data and the results of the data processing, where Figure 2(a) is the raw observation light curve data, and Figure 2(b) is the normalized light curve data.

3.2 Real-time Early Warning Model Structure and Algorithm Description

This section mainly introduces the GRU for the Warning of Abnormal Light Curve (GRUWALC) model of abnormal light curves based on a GRU algorithm and gives a detailed description of the algorithm. The network structure is based on the GRU model design, and the structure of GRUWALC is shown in Figure 3. The network structure includes an input layer, two GRU layers, a fully connected layer, and an output layer. The input data is the acquired light curve, and the output is the model to predict the output value of the light curve at the next moment. The specific parameter settings in the model are shown in Table 1. Among them, the first layer and the second layer represent GRU units, wherein the output dimension of the GRU of the first layer is 50, and the dimension of the GRU output of the second layer is 30. Then, the data output from the second layer is connected to the fully connected layer, the output dimension is 40, and the final output layer has a dimension of 1. In addition, other specific parameters are described as follows: the activation function uses tanh, and its function is to give the neural network model a nonlinear fitting ability. There are many activation functions, and tanh is used here; the proportional parameter dropout=0 for random killing of neuronal connections; and bias parameter b=0. Details of the algorithm are given in Algorithm 1.

The model input proposed in this paper is an 800-dimensional time series signal, which means that the first 800 data points are used as training samples. After the net-



(a) (b)

Fig. 2 Raw observation light curve data and normalized light curve data. (a) Raw observation light curve data. (b) Normalized light curve data.

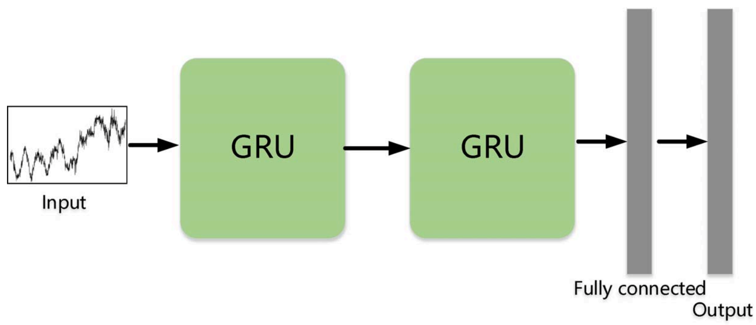


Fig. 3 GRUWALC network structure. Include input layer, two GRU layers, full connection layer and output layer.

Algorithm 1 The training process of GRUWALC

Input:
The number of training set, N_{train};
The number of test set, N_{test};
The number of samples taken from the training set for each training, $batchsize$;
The number of training times in the full sample of the training set, $epoch$;
Output:
prediction value, y_{pre};
for The loop count $< epoch$ **do**
  Choose $batchsize$ samples from training set.
  Save GRUWALC model.
  Save weights and parameters.
Test trained GRUWALC.
return y_{pre}.

Table 1 The Structure and Setting Parameters for GRUWALC Networks

No.	Layer	Output
1	GRU	50
2	GRU	30
3	Fully connected	40
4	Output	1

work model is trained, the model parameters are saved, and then the predicted values for the future period are given according to the trained model. Then the difference is calculated between the predicted value and the actual observed value, τ. If τ is greater than the set threshold ϵ, then we consider the observed value at this moment to be an abnormal value, where the threshold ϵ is defined as follows:

$$\epsilon = \Phi(|y_t - \hat{y}_t|)\,. \tag{4}$$

Here, y_t and $\hat{y}_t$ respectively represent the actual observations and prediction values of the model proposed in the training phase. When the model training converges, the stabilized function loss value is selected as the threshold, and the threshold is automatically driven by the data. In the process of model training, this paper uses Root Mean Square Error (RMSE) (Willmott & Matsuura 2005) as the loss function.

$$\text{Loss}_{\text{RMSE}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}\,. \tag{5}$$

Here, y_i is the real value, and $\hat{y_i}$ is the output of the GRUWALC model proposed in this paper. In this paper, we use the loss value that the model outputs when the training is stable as the threshold of the abnormal warning ϵ, and if this value is exceeded, an early warning is given.

4 EXPERIMENTAL ANALYSIS

This section mainly introduces the experimental results of the proposed algorithm and the comparison with the current anomaly detection algorithm, according to the description process of the Section 3 exception warning algorithm. First, the paper normalized the actual observation data, and then selected one of the collected samples as training data with a total dimension of 968. The first 800 data points are used as training samples, and the remaining observation data is used as test samples. Figure 4 presents the prediction results of the GRUWALC model proposed in this paper. The abscissa indicates the time of data acquisition, and the ordinate represents the result of normalization of the observed data. The blue line is the real data, and the orange line is the prediction result obtained by GRUWALC according to the training data. The number of data in the test data set is 168, and the gray area is the fluctuation range of the normal light curve. If the actual observation falls within the gray value range, it means that the observation data at this time are normal data, and if the actual observation is outside the gray range, the observation at this moment is considered to be an abnormal value and an abnormal warning is given. The results of partial amplification of the prediction data are shown in Figure 5. It can be seen from the experimental results that the algorithm proposed in this paper has a good prediction effect at most moments, but for times when the change is relatively severe, the prediction effect is not good, and a false alarm phenomenon occurs.

In this paper, because the number of samples is limited, we use a training set with a sample size of 800. In order to know whether different numbers of training samples will affect the prediction results, we set the number of training samples at 500, 600, 700 and 800 under a certain number of test sets, and then compared different RMSE values. The experimental results are shown in Figure 6. The experimental results show that the number of training sets is different and the RMSE values on the test set are also different. The reason for this difference is mainly due to the volatility of the observed data. For example, at 600 and 700 as the training set, the subsequent 168 test points are more volatile, so the calculated RMSE value is larger.

In order to compare the prediction performance of different algorithms on the light curve, the proposed GRUWALC algorithm is compared with other prediction algorithms, ARIMA, LSTM, LSTMCNN (Bartz et al.

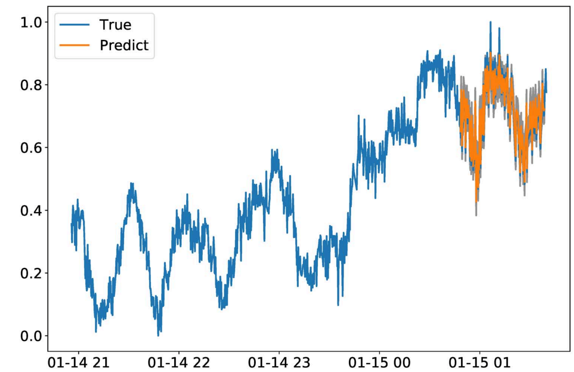


Fig. 4 GRUWALC model prediction results. The *blue line* is the real data, and the *orange line* shows the prediction results obtained by GRUWALC according to the training data. The *gray area* is the fluctuation range of the normal light curve data.

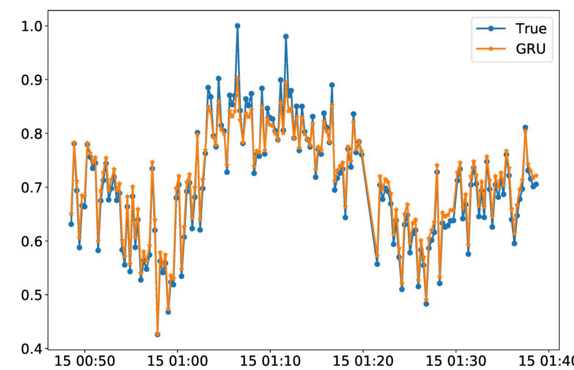


Fig. 5 Comparison of GRUWALC model predictions and real observations. The *blue line* is the real data, and the *orange line* shows the prediction results obtained by GRUWALC according to the training data.

2017), RNN and SimpleRNN (Elman 1990). In this section, the same number of training sets as GRUWALC are selected as training samples, and 168 data points are also selected as test sets. The prediction results of different algorithms are shown in Figure 7. The abscissa indicates the time at which the data was acquired, the ordinate indicates the normalized observation value. True indicates the actual observation. From the experimental results in Figure 7, the LSTM, LSMMCNN, RNN, SimpleRNN algorithms and the algorithm proposed in this paper can learn the nonlinear relationship of the data itself, so these algorithms have better prediction effects at most moments. However, for times when the change is relatively large, the prediction performance is not good. Since the ARIMA algorithm requires that the experimental data changes be stable or stable after the difference operation, it means that the relationship between the data is linear, and nonlinear data do not perform well, so the overall prediction effect is poor.

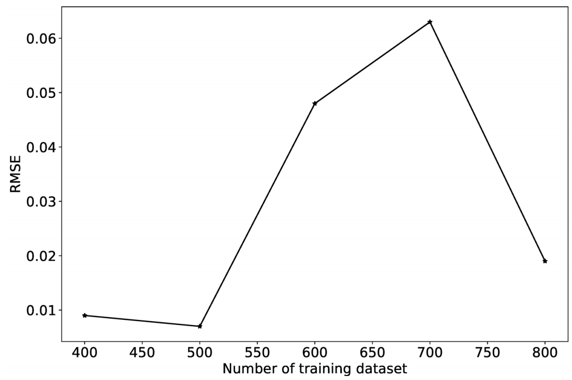


Fig. 6 Comparison of RMSE of different numbers of training samples. The abscissa indicates the different number of training samples. The ordinate indicates the RMSE value of the test set under different numbers of training samples.

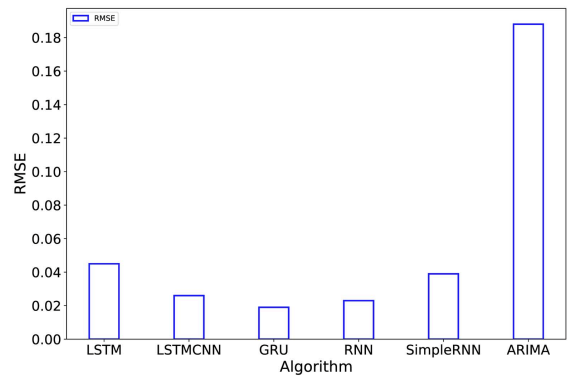


Fig. 8 RMSE comparison results of different algorithms. The abscissa represents different algorithms. The ordinate represents the RMSE of different algorithms.

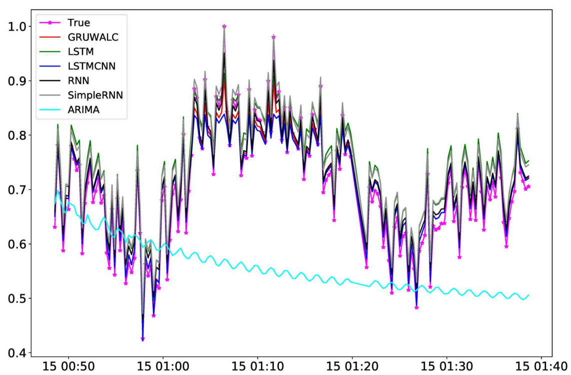


Fig. 7 Comparison of the prediction results of different algorithms. The abscissa indicates the time of data acquisition. The ordinate represents the normalized predicted value of different algorithms. True represents the actual observations.

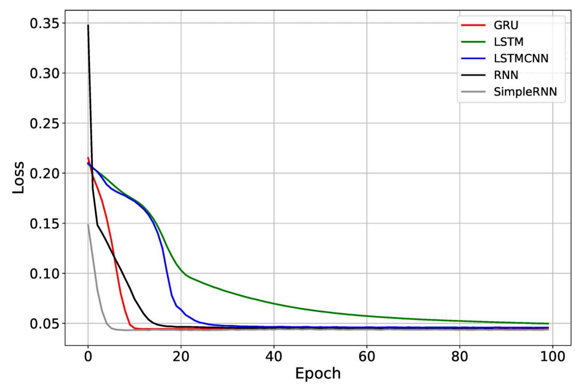


Fig. 9 Comparison of network training loss values. The abscissa represents the number of iterations. The ordinate indicates the loss value of the training phase.

This paper measures the predictive performance of different models by calculating the RMSE between the predicted and actual values of the model. This section uses RMSE as an indicator of whether a model fits the true data. If the RMSE value is smaller, the better the prediction performance of the model is. The RMSE comparison experiment results of different algorithms are shown in Figure 8. Figure 8 shows the prediction performance of different algorithms from the perspective of quantization. Among them, the algorithm based on a neural network is obviously superior to the traditional method of ARMIA. The GRUWALC algorithm proposed in this paper is more robust than other algorithms.

In practical applications, astronomers are also concerned about false alarms, that is, false alarm rate indicators (Scargle 1982). Traditionally defined anomaly methods use 3σ and 5σ criteria, where σ is the standard deviation. The formula is defined as follows:

$$\sigma(r) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - r)^2} \,. \tag{6}$$

Here, r is the mean of the training data set and x_i is the observed value of each sample point. We first calculated the standard deviation for the training set with a sample size of 800, then calculated 3σ and 5σ, and set 3σ and 5σ as thresholds respectively to get the false alarm rate for different algorithms. The final experimental results are shown in the last two rows of Table 2. From the experimental results, we know that in this experimental data, all the data of the test set exceeds 3σ and 5σ. Therefore, the anomaly detection method based on 3σ and 5σ as thresholds is very ineffective. The false alarm rate is within a certain time range, and the observed signal is not an abnormal signal. Due to the limitation of the automatic detection algorithm, the result is an abnormal signal. The false alarm rate for

Table 2 Comparison of False Alarm Rates of Different Algorithms

Threshold	**GRUWALC**	LSTM	LSTMCNN	SimpleRNN	RNN	ARIMA	0.6557	1.0928
0.0988	0	0	0.0007	0	0	0.0257	-	-
0.0368	0.0019	0.0519	0.0051	0.0396	0.0027	0.0488	-	-
0.0345	0.0027	0.0531	0.0055	0.0440	0.0079	0.0503	-	-
0.0307	0.0035	0.0551	0.0059	0.0500	0.0142	0.0523	-	-
0.0304	0.0043	0.0551	0.0063	0.0500	0.0146	0.0531	-	-
0.0279	0.0051	0.0567	0.0075	0.0535	0.0218	0.0551	-	-
0.6557	0	0	0	0	0	0	0.066667	-
1.0928	0	0	0	0	0	0	-	0.066667

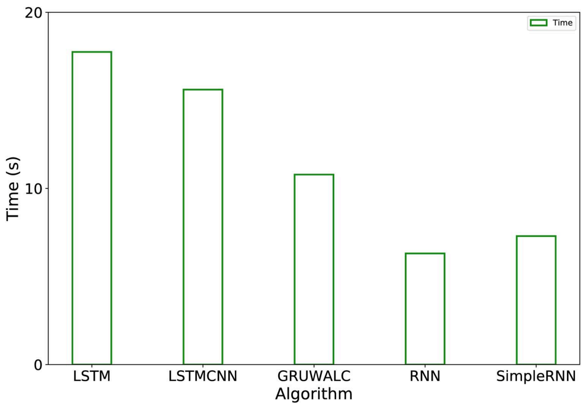


Fig. 10 Network training time comparison. The abscissa represents different algorithms. The ordinate represents the time it takes to iterate 100 times, in seconds.

the data used in this paper is defined as follows:

$$E = \frac{N}{nT}\,. \tag{7}$$

Here, E means the false alarm rate, n means the number of observation points, and T means the time interval between adjacent observation data points. According to Equation (4), we can find the number of false alarm exceptions in 168 observation points, and then calculate the false alarm rate according to Equation (7). The false alarm rates we get on the same test set according to different algorithms are shown in Table 2.

According to the false alarm rate results, the thresholds for anomaly detection we selected are different, and the false alarm rate is different. Generally, the larger the threshold, the lower the false alarm rate obtained on the normal light curve data set. Therefore, if the algorithm fits better, that is, the smaller the difference between the predicted value and the true value of the model, the false alarm rate is lower under the same abnormal threshold. According to the experimental results in Table 2, when the threshold is set to a small value, for example, the threshold is 0.0279, the experimental results show that the false alarm rate of the ARIMA algorithm is even smaller than that of the LSTM, but the prediction results of different algorithms (Fig. 7) and RMSE results (Fig. 8) show that the ARIMA prediction performance is far from the other algorithms. The reason for this phenomenon is that when the threshold is small, the absolute difference between the predicted value of the algorithm and the true value is greater than 0.0279 at most of the data points of the test set. So, at this time, the false alarm rate of most algorithms is very small. However, when the threshold increases, for example, when the threshold is 0.0988, the false alarm rate of the GRUWALC, LSTM, SimpleRNN and RNN algorithms is 0, while the false positive rate of ARIMA is large. The results of this experiment show that, if the threshold is 0.0988, the stronger the prediction ability of the model, the lower the false alarm rate.

In practical applications, the algorithm proposed in this paper does not require human intervention throughout the process, or requires only a small amount of human involvement. This frees astronomers from very time-consuming manual monitoring (Prabhu 2000) and allows them to concentrate more on the physical mechanisms involved. In addition, as the observed data is continuously updated, network parameters are adjusted based on new observations to ensure that they adapt to new data changes. Therefore, according to newly observed data, it is necessary to adjust the model parameters in time. For scenarios with high real-time requirements, the network convergence speed needs to be as fast as possible. This paper judges the network update speed by comparing the changes of the loss values of different algorithms during the training phase. The loss function value of different algorithms in the training phase is shown in Figure 9. The abscissa represents the number of iterations and the ordinate represents the loss value of the training phase. According to the change of the loss value in the training phase of the model, we find that the proposed GRUWALC and SimpleRNN algorithms converge faster, followed by LSTM and RNN. Since the LSTMCNN algorithm needs to learn more network parameters, its convergence speed is slower. In addition, the paper also records the time taken by different algorithms to reach stability during the training phase. The training time is compared as shown in Figure 10. The abscissa represents the different algorithms, and the ordinate represents the time required to iterate 100 times, in seconds. As mentioned in the real-time requirements discussed above, combined with the model prediction performance, this paper suggests GRUWALC as the prediction model, as it not only has a faster convergence speed (although the speed is not as fast as SimpleRNN) but can also meet the real-time requirements. Moreover, it has a lower RMSE, that is, model prediction is more robust.

5 CONCLUSIONS

In view of the high real-time light curve demands in time domain astronomy, this paper proposes a GRU-based real-time early warning model GRUWALC. This paper first establishes a prediction model for the time series data, and uses the historical data for the light curve to train the prediction model, saving the trained model and predicting the data of the next moment based on the trained model. When the RMSE of the actual observed value and the predicted value at the next moment exceed the set threshold, the observed value at this moment is considered to be an abnormal value and a real-time warning is given. It can be seen from the experimental results that the selection of the threshold has an important influence on the false alarm rate. Therefore, establishing a good predictive model can reduce the false alarm rate. The innovations of this paper mainly include two points. First, the GRU is introduced as a fitting model in the light curve prediction process for the first time; second, according to the established fitting model, the threshold for abnormal warning is set. By comparing the actual observation value and the model prediction value at the next moment, it is judged whether the observation value at the next moment is an abnormal value.

Acknowledgements This research was supported by the National Natural Science Foundation of China (No. 41672323) and the Interdiscipline Research Funds of Beijing Normal University. We are grateful for the data provided by the SVOM project team of the National Astronomical Observatories of the Chinese Academy of Sciences. We also thank the anonymous referee for a helpful report.

References

Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., et al. 2014, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22, 1533
Alard, C. 2000, A&AS, 144, 363
Bartz, C., Herold, T., Yang, H., & Meinel, C. 2017, in International Conference on Neural Information Processing, Springer, 880 (arXiv:1708.04811)
Battistelli, E. S., Amiri, M., Burger, B., et al. 2008, Journal of Low Temperature Physics, 151, 908
Bhatti, W. A., Richmond, M. W., Ford, H. C., & Petro, L. D. 2010, ApJS, 186, 233
Bramich, D. M. 2008, MNRAS, 386, L77
Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. 2014, arXiv:1409.1259
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Dalcanton, J. J., Canizares, C. R., Granados, A., Steidel, C. C., & Stocke, J. T. 1994, ApJ, 424, 550
Dickey, D. A. 1984, Biometrika, 71, 599
Elman, J. L. 1990, Cognitive science, 14, 179
Graham, M. J., Djorgovski, S. G., Mahabal, A., et al. 2012, Distributed and Parallel Databases, 30, 371
Hinton, G., Deng, L., Yu, D., et al. 2012, IEEE Signal Processing Magazine, 29, 82
Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735
Hon, M., Stello, D., & Yu, J. 2017, MNRAS, 469, 4578
Jones, O. C., Woods, P. M., Kemper, F., et al. 2017, MNRAS, 470, 3250
Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436
Liu, W., Zhu, M., Dai, C., et al. 2019, MNRAS, 483, 4774
Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. 2018, Nature Astronomy, 2, 151
Prabhu, T. P. 2000, Bulletin of the Astronomical Society of India, 28, 233
Sarikaya, R., Hinton, G. E., & Deoras, A. 2014, IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 22, 778
Scargle, J. D. 1982, ApJ, 263, 835
Sguera, V., Bazzano, A., Bird, A. J., et al. 2006, ApJ, 646, 452
Telezhinsky, I., Eckert, D., Savchenko, V., et al. 2010, A&A, 522, A68
Willmott, C., & Matsuura, K. 2005, Climate Research, 30, 79
Young, T., Hazarika, D., Poria, S., & Cambria, E. 2018, ieee Computational Intelligence Magazine, 13, 55 (arXiv:1708.02709)
Zaremba, W., Sutskever, I., & Vinyals, O. 2014, arXiv:1409.2329
Zhang, G.-R., Yin, K.-L., Liu, L.-L., Xie, J.-M., & Du, H.-L. 2005, Yantu Lixue (Rock Soil Mech.), 26, 1312
Zhang, R., Lin, L., Zhang, R., Zuo, W., & Zhang, L. 2015, IEEE Transactions on Image Processing, 24, 4766
Zhao, W., & Du, S. 2016, IEEE Transactions on Geoscience and Remote Sensing, 54, 4544