Short-Timescale Gravitational Microlensing Events Prediction with ARIMA-LSTM and ARIMA-GRU Hybrid Model

Ying Sun, Zijun Zhao, Xiaobin Ma, and Zhihui Du(✉)

Department of Computer Science and Technology, Tsinghua University, Beijing, China
duzh@tsinghua.edu.cn

Abstract. Astronomers hope to give early warnings based on light-detection data when some celestial bodies may behave abnormal in the near future, which provides a new method to detect low-mass, free-floating planets. In particular, to search short-timescale microlensing (ML) events from high-cadence and wide-field survey in real time, we combined ARIMA with LSTM and GRU recurrent neural networks (RNN) to monitor all the observed light curves and to alert before abnormal deviation. Using the good linear fitting ability of ARIMA and the strong nonlinear mapping ability of LSTM and GRU, we can form an efficient method better than single RNN network on accuracy, time consuming and computing complexity. ARIMA can reach smaller alerting time and operating time, yet costing high false prediction rate. By sacrificing 15% operating time, hybrid models of ARIMA and LSTM or GRU can achieve improved 14.5% and 13.2% accuracy, respectively. Our work also provide contrast on LSTM and GRU, while the first type is commonly used for time series predicting systems, the latter is more novel. We proved that in the case of abnormal detection of light curves, GRU can be more suitable to apply to as it is less time consuming by 8% while yielding similar results as LSTM. We can draw a conclusion that in the case for short-timescale gravitational microlensing events prediction, hybrid models of ARIMA-LSTM and ARIMA-GRU perform better than separate models. If we concentrate more on accuracy, ARIMA-LSTM is the best option; on the other hand, if we concentrate more on time consuming, ARIMA-GRU can save more time.

Keywords: Gravitational lensing · Recurrent neural networks · ARIMA · Time series prediction and alarming

1 Introduction

Astronomy is the origin of information explosion, and it is the first field to meet the challenge of big data [1]. In the 21st century, astronomical data is growing

J. Li et al. (Eds.): BigSDM 2018, LNCS 11473, pp. 224–238, 2019.
https://doi.org/10.1007/978-3-030-28061-1_23

at a rate of terabytes or even PB. By 2010, the information file had been as high as 1.4×242 bytes. The ground-based Wide-angle Camera array (GWAC) [15] is part of the SVOM space project, which searches for various types of optical transient sources by continuously imaging the 5000 square-degree field of view (FOV) every 15 s. Each exposure contains $36 \times 4\text{k} \times 4\text{k}$ pixels, usually 36×175, 600 sources can be extracted. A GWAC camera produces a 32-megapixel map every 15 s. According to the limit star 16.5, which produces 2,400 images per night, each night the GWAC project will generate 1010 recorded star catalog data. The data rate per camera is approximately 12,000 points per second (2.4 MB), which means the total data rate for the entire GWAC camera array system is 85 MB/s. The image data has about 1.7*105 records per image. The entire camera array produces 6.12*106 records in 15 s. Each night is observed for 10 h, and each record has 22 columns of attributes. GWAC will generate about every day. With 2.5 TB of photometric star table data, the 10-year design cycle will form a very large-scale database with a total number of stars with a scale of 3 PB 6 PB, which requires the database management system to have a very strong processing capacity for massive data. The China GWAC Observation Core Station is located in the Xinglong Observation Base of Hebei Province. GWAC's international station in Chile is under discussion. Xinglong Station will have 9 GWAC turntables and 36 CCD cameras of 18 cm [15]. In addition, the core station has a set of mini-GWAC system, which is the leading project of the ground-based wide-angle camera array. It consists of 12 sets of 8 cm large field of view cameras and has been built and placed in the Xinglong Observatory of the National Astronomical Observatory. In this paper, we rely on the mini-GWAC data as samples of training and testing. We investigate the problem of a real-time search for short-timescale gravitational ML events from a huge number of light curves by applying hybrid model of ARIMA-LSTM and ARIMA-GRU. Experiments are conducted on mini-GWAC dataset to evaluate the performance of these two models.

2 Related Work

Impressive works have been conducted in the area of short-timescale gravitational microlensing events prediction. Early Warning System (EWS) was first developed by Udalski in 1994, which is a real-time search system that has succeeded in operation in OGLE project for more than 20 years [7,8]. In 2006, Wyrzykowski et al. [6] have succeeded in systematic search for short-timescale gravitational microlensing events from the variable baselines in OGLE-III data. Recently, an approach of ARIMA that increases efficiency with a multi pro- cess parallel approach was developed by Bi [12]. The ARIMA model is improved with dynamic and parallel processing to detect anomaly. Also, a vari- ant form of ARIMA called D-ARIMA is developed to adjust the parameters of ARIMA for real-time anomaly detection on light curves [9]. Additionally, we look into algorithms of time series detection based on neural networks. In 1987, Farmer and Sidorowich [11] have yield robust results by applying neural network to predict chaotic time series, reaching high effi- ciency in Mackey-Glass delay differential equation, Rayleigh-Benard convection, and Taylor-Couette flow. Later, the advantages of recurrent neural networks are discovered and one of its variant, LSTM, becomes most widely applied [10]. In addition to LSTM, A gated recurrent unit (GRU) was proposed by Cho et al. [13] to make each recurrent unit to adaptively capture dependencies of different time scales, yielding similar results as LSTM yet less computing complexity.

3 ARIMA-LSTM and ARIMA-GRU Hybrid Models

Light-detection data from mini-GWAC implies over 900 time series files for each planet. It is obvious that data is assumed to be consisted of both linear part and the nonlinear part [2]. Thus, we can express the data as follows.

$$x_t = L_t + N_t + \epsilon_t$$

In the function of data x_t, L_t represents the linearity of data at time t, while N_t signifies nonlinearity. The ϵ_t value is the error term. In previous work [3], the Autoregressive Integrated Moving Average (ARIMA) has a great performance on linear problems of over 85% accuracy in general. It is a traditional method in time series prediction. On the other hand, the Long Short-Term Memory (LSTM) model can capture nonlinear trends in the dataset. As a result, we develop one hybrid ARIMA-LSTM to encompass both linear and nonlinear tendencies in the model.

3.1 ARIMA+LSTM Model

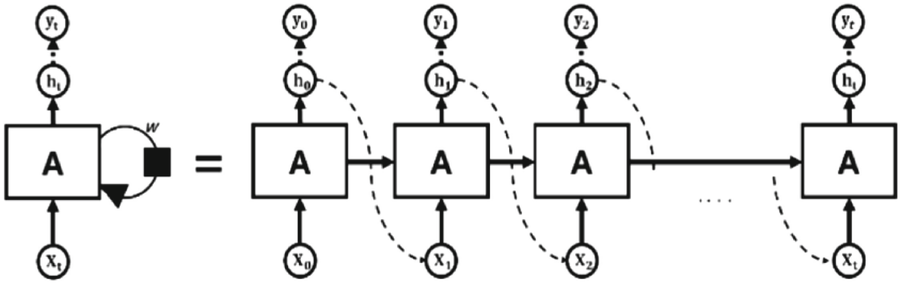


Fig. 1. Structure of recurrent neural network

As we know, ARIMA is a kind of traditional model used to predict the results and LSTM is a kind of neueal network. Our hybrid model which combines ARIMA and LSTM model to improve precision (Fig. 1).

The meaning of ARIMA is the one-product autoregressive moving average process. Assuming that a stochastic process contains d unit roots, which can be transformed into a stationary autoregressive moving average process after d difference, the stochastic process is called integral autoregressive moving average process. In this hybrid model, ARIMA is used for linear part of prediction.

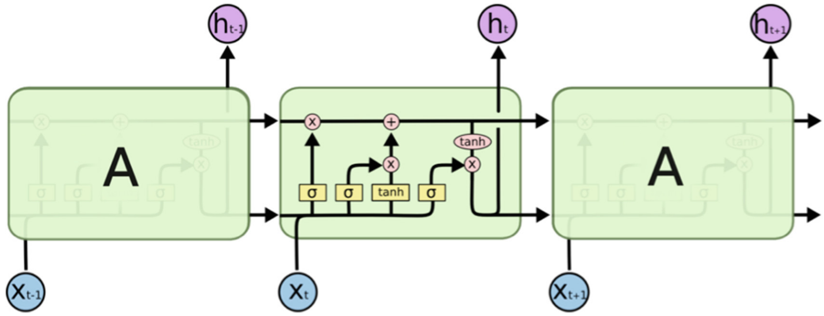


Fig. 2. Structure of LSTM

LSTM is a special type of RNN that can learn long-term dependency information, as shown in Fig. 2. LSTM was proposed by Hochreiter and Schmidhuber [4] and has recently been improved and promoted by Graves [5]. In many problems, LSTM has achieved considerable success and has been widely used. In this hybrid model, LSTM is used to predict nonlinear part.

In terms of final predictions, the first half of each file from the dataset is used as testing sets and the second half of each file from the dataset is used as training sets.

(1) In the linear predictions, training sets can be used directly to predict the future. In each file, the first 20% of the series is used as the basis and this is the size of the moving window. When a new result is predicted, the new result will add in and the whole window will move one step forward until the end. At the same time, the residuals, which means the training sets for LSTM, are also collected.
(2) In the nonlinear predictions, testing sets are from the residuals produced by the linear predictions. And the way how to train the neural network is of significance. Each training set is divided into two parts. The first part which is the first 20% of this series is as the training sets for LSTM and the rest is the test sets to study in order to help LSTM improve precision. As a result, the nonliear predictions are produced.
(3) Finally, predictions consist of linear part and nonliear part.

There are a lot of advantages of this hybrid model. Compared to pure ARIMA model, this model can produce preciser results with sacrificing little time. And compared with pure neural network, it is easier for user to adjust the parameters in a quicker and proper way. However, there are some challenges this hybrid model needs to face, such as reducing time cost and so on.

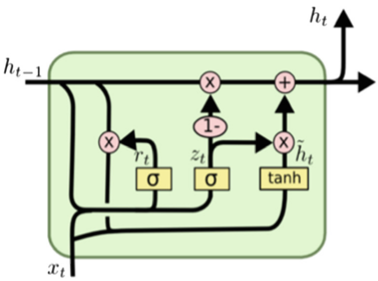
3.2 ARIMA+GRU Model

This model is designed to compare the efficiency of hybrid model of ARIMA and LSTM. We replace LSTM with GRU as GRU was widely applied since 2015 and it can yield better results in some situation than LSTM. Datasets of Short-timescale Gravitational Microlensing Event are used to evaluate whether hybrid models of GRU is more efficient in this background.

As a variant of LSTM, GRU combines the forgotten gate and the input gate into a single update gate, as shown in Fig. 3. The cell state and the hidden state are also mixed, with some other changes. The final model is simpler than the standard LSTM model and is a very popular variant. The construction of the GRU is simpler: one gate less than the LSTM, so there are fewer matrix multiplications. As a result, GRU can save a lot of time when the training data is large, which is also proved to be true in our case.

And the difference between this model and the previous hybrid model combined with LSTM is only the neural network. All the steps are identical.

And the parameters used in neural network are listed below.



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Fig. 3. Structure of GRU

Table 1. Sizes of the models tested in the experiments

Algorithms	Units	Parameters
LSTM	128	70400
GRU	128	56192

4 Experiment Method

4.1 Dataset

The data of GWAC has not been open till recently, so our algorithms are tested on mini-GWAC dataset. Mini-GWAC system is consisted of 12 sets of 8 cm large field of view cameras and has been built and placed in the Xinglong Observatory of the National Astronomical Observatory. In this paper, we rely on the mini-GWAC data as samples of training and testing. For each planet, mini-GWAC dataset contains average of 980 txt files, each indicating a thorough observation of one single approach. In each file, we can receive 900 data consisting a part of time series. Contrast of main station of GWAC and mini-GWAC is presented in Table 2.

The experimental environment is based on CentOS Linux and python 3.6.

Table 2. Comparison of GWAC and mini-GWAC.

GWAC station	Main station	Mini-GWAC
Numbers of turrets	9	6
Numbers of cameras	36 cameras in total and 4/turret	12 cameras in total and 2/turret
Diameter	18 cm	7 cm
Focal length	213 mm	85 mm
Length of wave	500–850 nm	500–850 nm
Limiting magnitude	16.5 V	13.0 V
Binocular FOV	4k*4k in pixel;12.8*12.8 = 160	12.8*12.8 = 16
Monocular FOV	5000°	5000°
Pixels	12	2

4.2 Model Fitting

Before fitting the ARIMA model, the order of the model must be specified. The ACF plot and the PACF plot can be used to aid the decision process. ACF/PACF plots of the light-curve data for star11 is plotted in Fig. 4.

In our ARIMA algorithm, we set (p,d,q) = (0,1,1) because it can yield the best outcome of Akaike Information Criterion (AIC) [14]. We compute the log likelihood function for the AIC metric by using the maximum likelihood estimator. We fit the ARIMA model with data from one planet at a time. For each file, we assume that the first 50 data is normal and learn 51th out of them. Then, we move forward at step = 1 at a time until we go through the whole file. After fitting the ARIMA model, we generate predictions of the linear part of the data. Then, we feed the residence of prediction results and reality value into LSTM and GRU networks for training and testing. For each individual planet, we train all previous samples and test light-curve data of the latest day to measure whether there is short-timescale gravitational microlensing events. LSTM and GRU are both imported from keras in python, using 128 units. Final results are

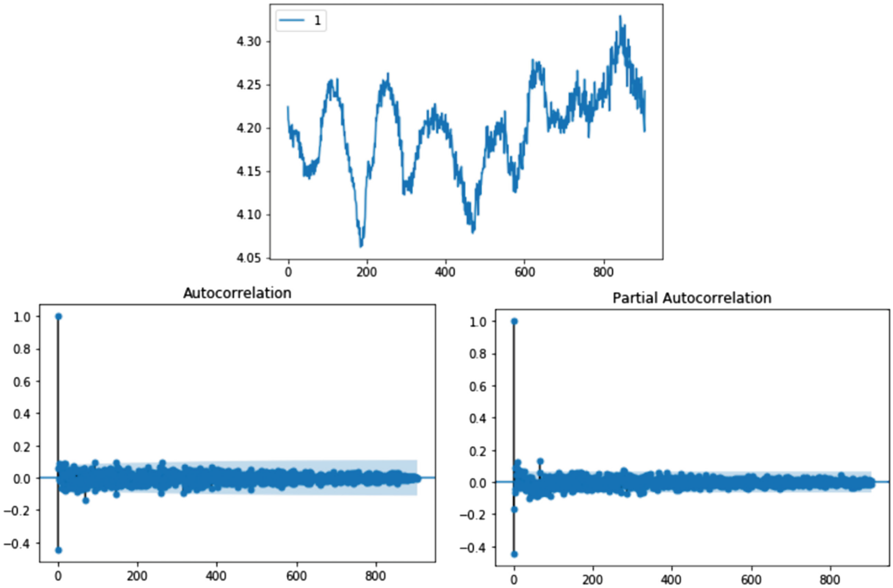


Fig. 4. Notable trends in the data and its ACF/PACF

the sum of the outputs of ARIMA and LSTM or GRU, forming ARIMA-LSTM and ARIMA-GRU hybrid model. An alarm will be set out if the difference of prediction and reality reaches a threshold. To determine the value of threshold, we tested several value shown in Table 3. The threshold is 0.2 eventually.

Table 3. Evaluation results with different thresholds

Thresholds	Loss of epoch	Alerting time
0.1	e−5	−37.5%
0.15	e−5	42.6%
0.2	e−5	42.2%
0.25	e−5	45.4%
0.35	e−5	47.5%

4.3 The Algorithm

Algorithm 1. ARIMA-LSTM/ARIMA-GRU Hybrid Model

$Result = []$
$Resid = emptylist$
for *dataset* in *datasets* do **do**
  $Predict = []$
  $Models = emptylist$
  $Order = arima.aicminorder$
  if unstable: **then**
    $Model = diff(model)$
  end if
  $Model = fitarima(dataset, order)$
  $Addresidual[0]toResid$
  $Addpredict[0]toPredict$
  $AddpredicttoResult$
end for
$SaveResult[-1], Resid$

5 Results and Evaluation

5.1 Testing Results on Simulation Dataset

Before testing on mini-GWAC dataset, we first test on a simulation dataset for evaluation under GWAC-like environment was generated, including 12,960 constant sample light curves and 3240 variable sample light curves. We first combine the microlensing magnification and baseline by three steps: (1) zero-pad the microlensing magnification at the right to expand its duration from T' days to 3+T' days; (2) element-wisely add IA and Is0; and (3) horizontally reverse the combined signal.

To meet the convention term of signal processing, source brightness lensed part in light curve (time length is T) will be called the signal part, while the other part is defined by the background (no lensing effect) in our paper. In our simulation, we will generate a background light curve with a time span of 3 + T' days.

We measured time consuming and accuracy of our hybrid model on simulated dataset. The dataset includes 12,960 constant sample light curves and 3240 variable sample light curves, divided into 636 different files. In this section, the ratio of training set to testing set is 5:1.

In each testing sample, each file is consisted of roughly 1,200 data. After training LSTM or GRU model with training set, we further use the first quarter of each sample, about 300 data in advance, for testing. The line of reality as well as prediction are drawn in Figs. 5, 6, 7, 8 and Fig. 9.

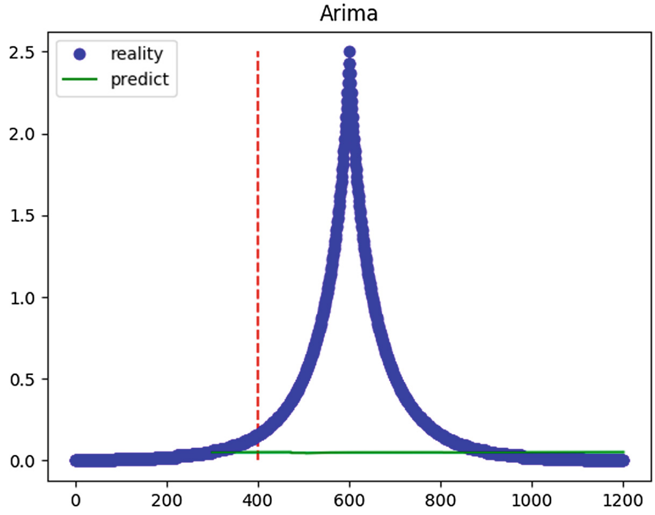


Fig. 5. ARIMA test on simulated dataset

The total time of evaluating these 16,200 sample stars by hybrid model is about 8 h and 24 min, which means that the model is capable to handle at least 16,200 stars in one day and can be applied to online daily renewable systems.

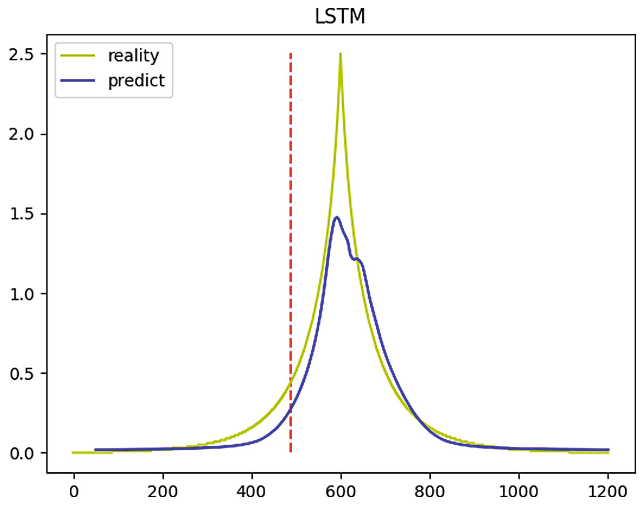


Fig. 6. LSTM test on simulated dataset

Results are drawn as follows:

(1) The average time consuming of one star is less than 2 s.
(2) Accuracy of hybrid models can reach over 90%, yet the time of alerting differs from one another.

Table 4. Evaluation results of different algorithms

Algorithms	Accuracy/Alerting time	Execution time
ARIMA	84.375%/40.0%	1.84 s/star
LSTM	88.67%/51.2%	0.56 s/star
GRU	87.33%/52.3%	0.54 s/star
ARIMA-LSTM	98.31%/40.2%	1.85 s/star
ARIMA-GRU	98.30%/41.4%	1.82 s/star

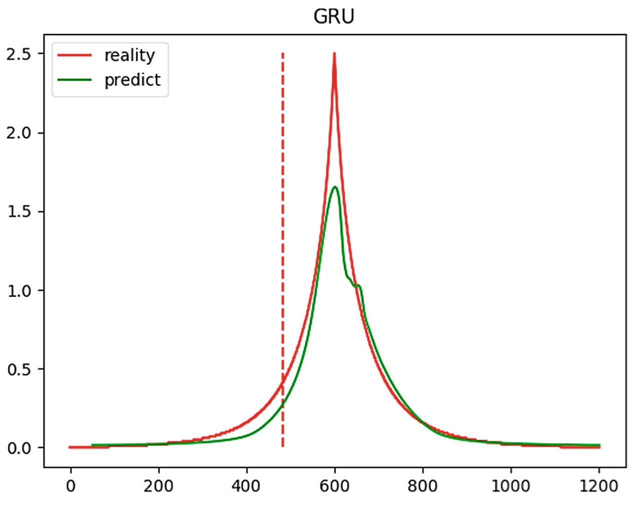


Fig. 7. GRU test on simulated dataset

5.2 Testing Results on Mini-GWAC

In this section, we measured three aspects of the efficiency of the model - accuracy, time consuming and computing complexity. Accuracy is determined by the percentage of alerting point and false prediction rate. That is, we hope the model should be precise on predicting and sensitive on abnormal cases at the same time. We marked the point where the alarm is set out in red, as shown in

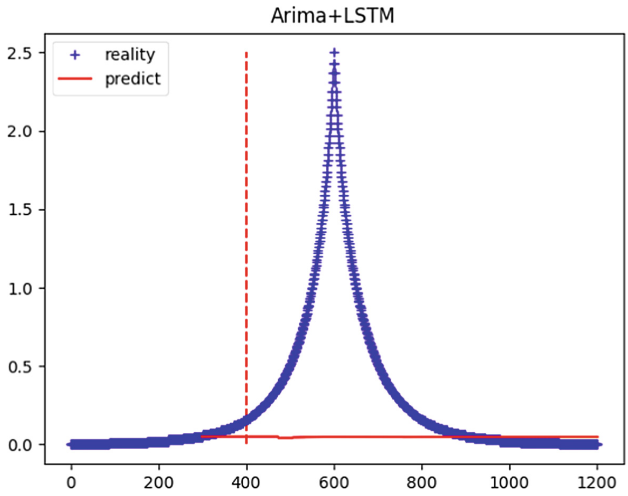


Fig. 8. LSTM-ARIMA test on simulated dataset

Fig. 10. Alarms should be set out before 50% where the abnormal series reaches its peak. The smaller alerting time is, the better the model performs. False prediction is the offset rate of the prediction value and reality value.

Additionally, we measure time consuming by operating time. Combined with different structure of algorithms in Sect. 4.3, it can also indicate computing complexity. Operating time is the prediction time for the single latest observation file. Results can be drawn from Table 4 for mini-GWAC dataset:

(1) GRU trains faster than LSTM for short-timescale gravitational microlensing events prediction.
(2) Alarm timeliness differs little among tested methods. The accuracy is slightly better for hybrid models. LSTM behaves more robust than GRU.
(3) GRUs are simpler and thus easier to modify, for example adding new gates in case of additional input to the network. This results in less training time and computing complexity.
(4) ARIMA can reach smaller alerting time and operating time, yet costing high false prediction rate. By sacrificing 15% operating time, hybrid models of ARIMA and LSTM or GRU can achieve improved 14.5% and 13.2% accuracy respectively (Table 5).

Table 5. Evaluation results of different algorithms

Algorithms	Accuracy/Alerting time	Execution time
ARIMA	81.60%/41.7%	0.349 s/star
LSTM	93.72%/42.6%	0.478 s/star
GRU	93.28%/43.3%	0.440 s/star
ARIMA-LSTM	96.11%/42.2%	0.406 s/star
ARIMA-GRU	94.83%/42.8%	0.413 s/star

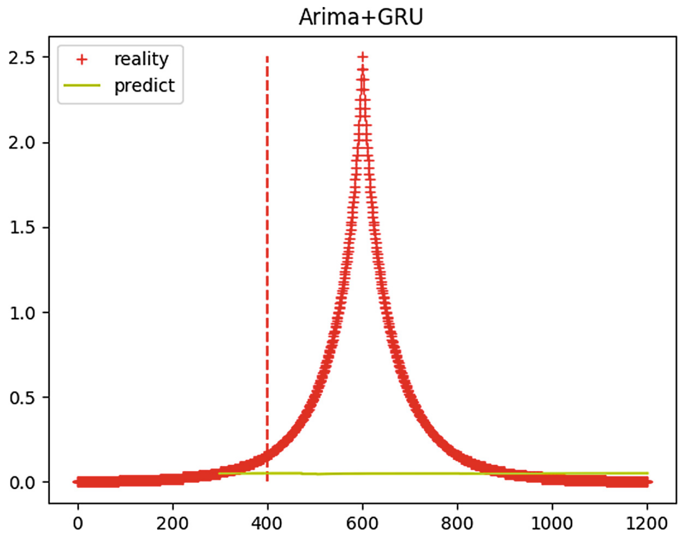


Fig. 9. GRU-ARIMA test on simulated dataset

5.3 Comparison Between LSTM and GRU

The key difference between a GRU and an LSTM is that a GRU has two gates (reset and update gates) whereas an LSTM has three gates (input, output and forget gates). The GRU unit controls the flow of information like the LSTM unit, but without having to use a memory unit. It just exposes the full hidden content without any control. In the algorithm we apply to mini-GWAC dataset, the number of units are set to be 128. See Table 1 for the details of the model sizes.

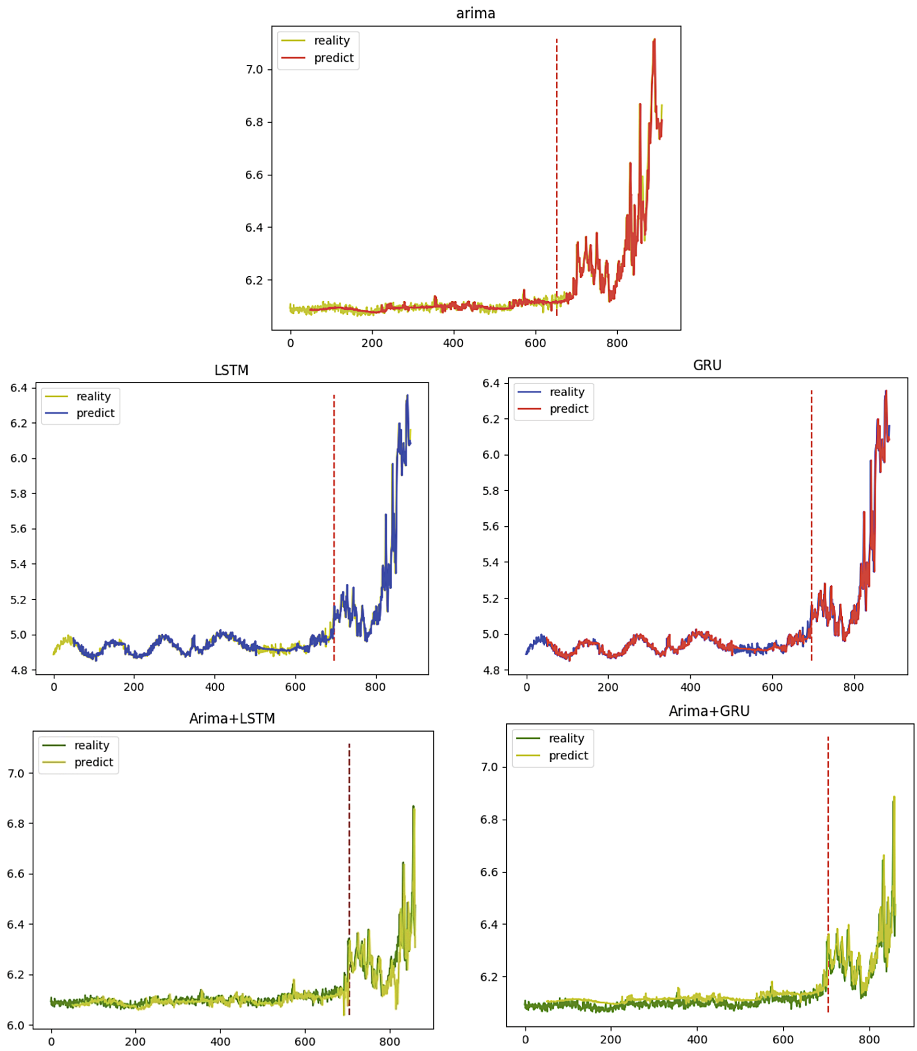


Fig. 10. Test on Star11 of different algorithms

6 Conclusion

Given by the significant role of abnormal activities like micro-lensing for asteroid, computer-aided forecasting is a heavily studied problem in the area of early detection. Like many other studies of this issue, time series are trained and tested by applying neural network which can yield high accuracy of over 90%. However, in the case of mini-GWAC dataset, or for future GWAC dataset, time consuming and computing complexity should be considered as more important criterion for a model. This paper proposed an algorithm of hybrid ARIMA-LSTM and ARIMA-GRU to replace single RNN approach:

ARIMA can reach smaller alerting time and operating time, yet costing high false prediction rate. By sacrificing 15% operating time, hybrid models of ARIMA and LSTM or GRU can achieve improved 14.5% and 13.2% accuracy respectively. Proposed hybrid methods perform better than ARIMA and RNN in the case of mini-GWAC short-timescale gravitational microlensing events prediction.

Additionally, it provides difference between these two methods to indicate strength and weakness of LSTM and GRU which both generate from RNN. This can help with future work of applying different methods to different backgrounds and datasets. Overall, we argue that the proposed work can provide as an essential approach in computer-aided forecasting of time series. Future work contains improvement on strengthening the model and its performance, an experiment test on GWAC dataset, minimizing time consuming especially in the process of neural network training.

Acknowledgement. This research is supported in part by Key Research and Development Program of China (No. 2016YFB1000602), "the Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China", National Natural Science Foundation of China (Nos. 61440057, 61272087, 61363019 and 61073008, 11690023), MOE research center for online education foundation (No 2016ZD302).

References

1. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution that Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, Boston (2013)
2. Zhang, G.P.: Time series forecasting using a hybrid arima and neural network model. Neurocomputing **50**, 159–175 (2003)
3. Konar, A., Bhattacharya, D.: An introduction to time-series prediction. Time-Series Prediction and Applications. ISRL, vol. 127, pp. 1–37. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54597-4_1
4. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). Massachusetts Institute of Technology
5. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
6. Wyrzykowski, Ł., et al.: OGLE-III microlensing events and the structure of the Galactic bulge. The American Astronomical Society (2015)
7. Udalski, A.: AcA **53**, 291 (2003)
8. Udalski, A., Szymanski, M.K., Szymanski, G.: AcA **65**, 1 (2015)
9. Feng, T., Du, Z., Sun, Y., et al.: In 6th 2017 IEEE International Congress on Big Data (Honolulu, HI), p. 224 (2017)
10. Lipton, Z.C., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning. arXiv. 1506.00019
11. Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. Phys. Rev. Lett. **59**, 845 (1987)

12. Bi, J., Feng, T., Yuan, H.: Real-time and short-term anomaly detection for GWAC light curves. Comput. Ind. **97**, 76–84 (2018)
13. Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014)
14. http://www.statisticshowto.com/akaikes-information-criterion/
15. Wan, M.: An application research of column store MonetDB database on GWAC large-scale astronomical data management. National Astronomical Observatories, Chinese Academy of Sciences (2016)