Detection of GWAC Abnormal Light Transform Based on Sparse Autoencoder

M. Zhu, X. C. Yu, F. Q. Duan

Beijing Normal University

Introduction

The Geographic Wide Angle Camera Array (GWAC) has a large field of visual (5000 square degrees) and high-speed sampling (15 seconds), so GWAC has greater ability of monitoring function of area and higher resolution in time-domain astronomy. However, in the post processing stage of GWAC data, there are many difficulties in data mining, such as abnormal detection of light curves. The detection of abnormal light curves from large amounts of data has caused astronomers to be in a state of fatigue for a long time, thus increasing the probability of misjudgement and abnormal detection. However, GWAC has put forward higher requirements for the efficiency and accuracy of the abnormal detection of the light curves.

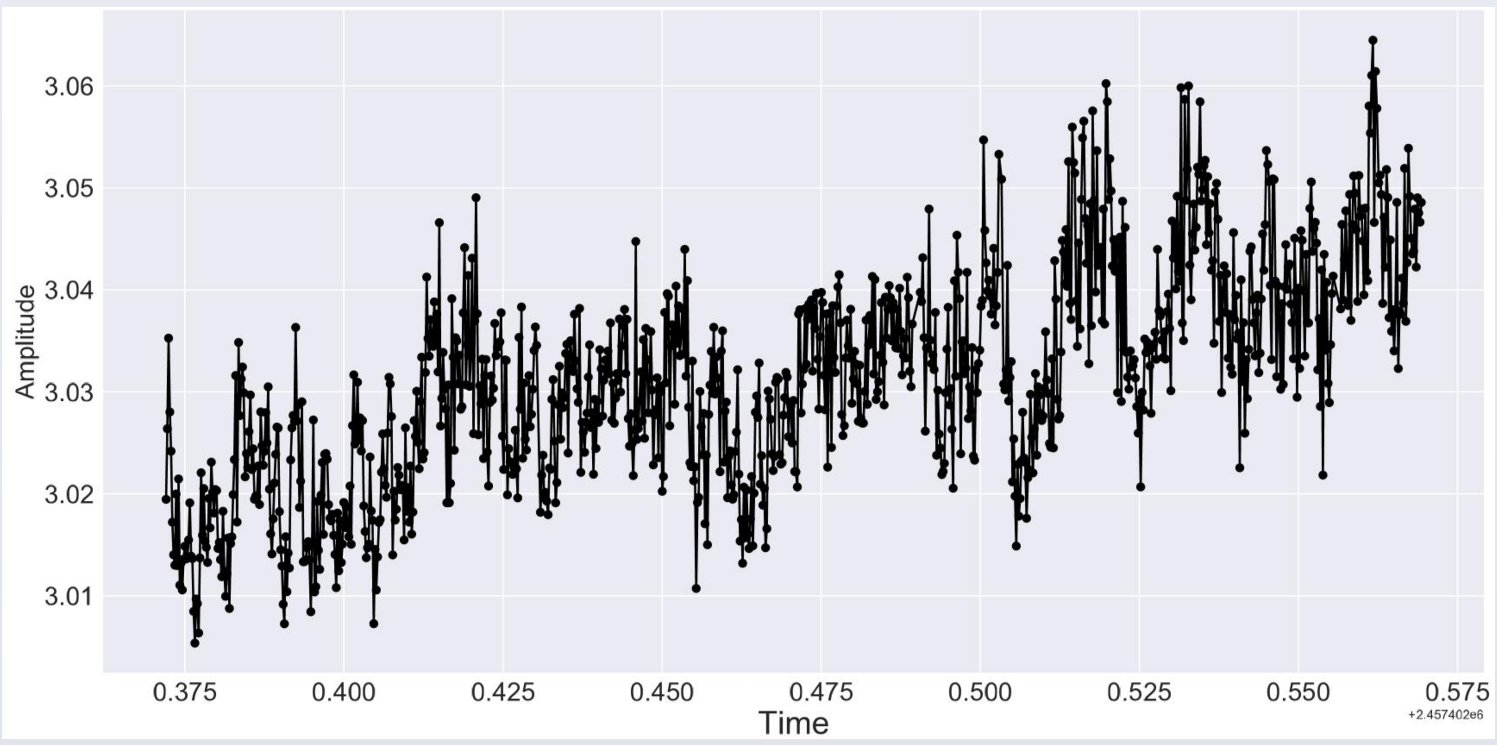

Fig.1. Light curve. Each light curve contains 900 observations and the time interval between neighbor observation points is 15 seconds

.

Method

In this paper, we firstly make the basic framework for SAE and generate the training samples in form of a sliding window. Secondly, we use normal samples and Back Propagation (BP) method to train and optimize the whole network model to get the corresponding error distribution threshold of normal samples. Finally, according to the reconstruction error of the test samples, we detect typical abnormity in the light curves. Experiments show that our method can get threshold of the reconstruction error distribution of normal samples in case of sample imbalance. Meanwhile, it can detect abnormity of light curves accurately.

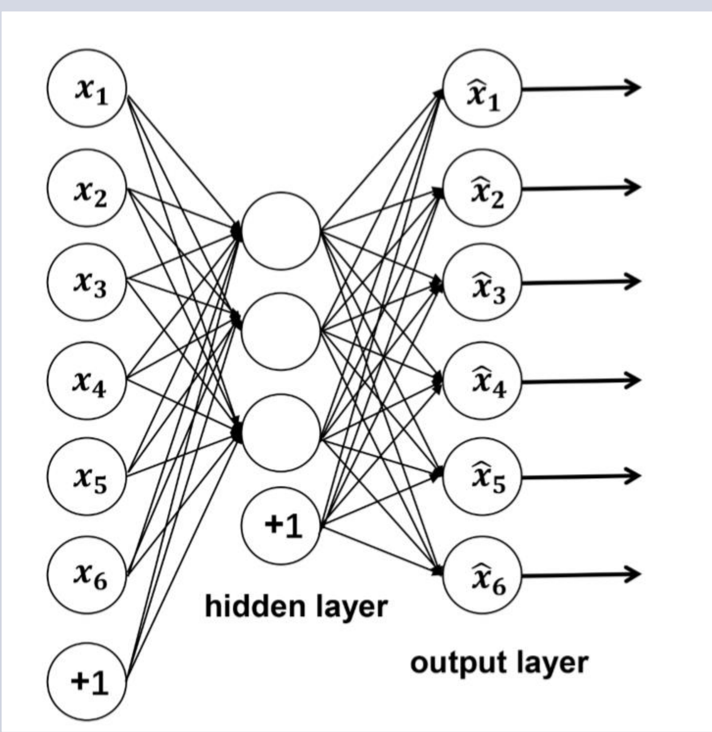

Figure.2. The structure of an sparse autoencoder. The framework can be considered to consist of the encoder and the decoder.

Results

In our work, normal samples are extracted from 300 light curves for training SAE to obtain the reconstruction error distribution threshold of normal data. Finally, the trained SAE is applied to detect the data of abnormal light curves. Therefore, when there are outlier, the value of cost function surpass the range of threshold. Therefore, if the threshold value is exceeded, the data of the light curves data at that time can be identified to be abnormal. In addition, all light curves use 0 to indicate the normal data at that time and 1 is abnormal.

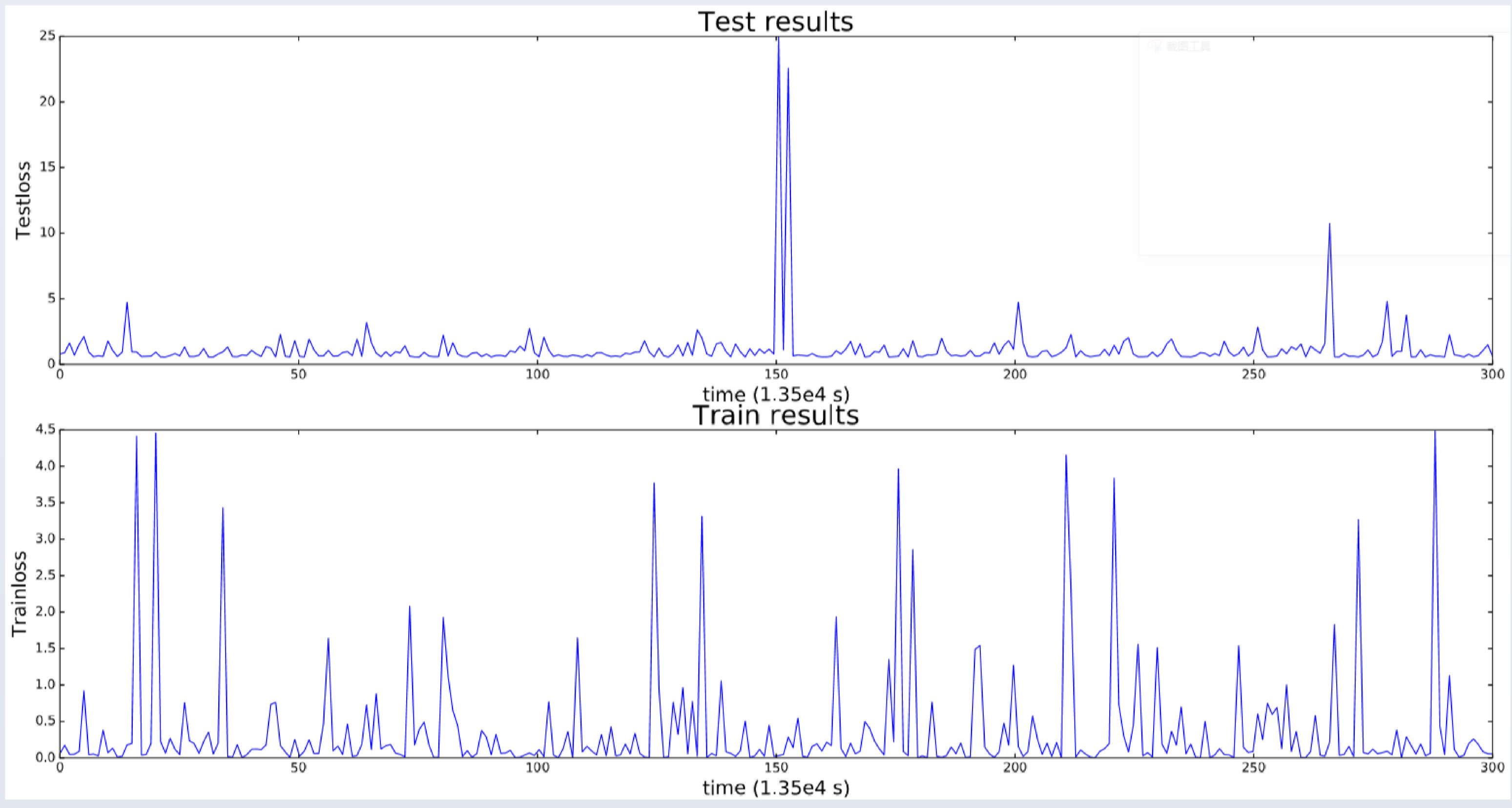

Figure.3. Results of SAE tests on light curves. With the maximum train loss of training data (bottom) as the threshold (4.48), 6 outliers (top) are detected in the test stage.

Algorithm	False positive error (%)	Elapsed Time (s)
SAE	0.50	2.27
SAE4	0.25	1.77
SAE2	0.00	1.67
K-means	0.96	1.10

Conclusion

In this paper, we import an algorithm based on unsupervised mechanism that enables us to discover outliers in light curves. To do so, we train an SAE and obtain an approximation to the identity function, which is used for determining our outlier score. However, our method is just time-consuming in the training and is fast in the unknown light curve analysis, allowing astronomers to explore very large data sets. Moreover, our approach is not limited to astronomical problems and can be applied to any database in which anomaly detection is necessary. On the other hand, we are planning to apply our algorithm to more observations. Through this analysis, we hope that we could find more unknown transient objects.

Reference

Xiong, Liang, Poczos, Barnabas, Connolly, Andrew, Schneider, Jeff, 2014, Available: http://www.ml.cmu.edu Landolt A. U., 1990, PASP, 102, 1382

Protopapas P., Giammarco J. M., Faccioli L.et al., 20061, MNRAS, 369, 677

Belokurov V., Evans N. W., Du Y.L., 2006, MNRAS, 341, 1373

Belokurov V., Evans N. W., Du Y.L., 2006, MNRAS, 352, 1365