



Anomaly Detection Method Based on Granger Causality Modeling

Siya Chen^(✉), G. Jin, Sun Peng, and Lulu Zhang

National University of Defense Technology, Changsha, China
2280526009@qq.com

Abstract. Satellites are very expensive to manufacture and require high reliability. Monitoring a large amount of telemetry data during the satellite orbit operation, the telemetry data are an important data source for analyzing the internal correlation of the satellite system and detecting anomalies. Telemetry data is in the form of time series, and there may be mutual influence and correlation between these time series. Due to the diversity of its influence and association forms, it is necessary to establish an effective model to determine the association relationship between them in order to detect anomalies on this basis and identify the cause of anomalies. In this paper, we use Granger causality model to analyze correlation between time series of telemetry data and establish a causality model. Detecting anomalies according to the causality which under normal circumstances and find out the cause of the anomalies. The case study shows that our method is effective.

Keywords: Granger analysis · Causal relationship · Anomaly detection · Satellites

1 Introduction

There has been a lot of work on association analysis for time series. Most researches use time series correlation coefficient to determine whether there is a correlation. For example, Pearson correlation coefficient [3] and Spearman correlation coefficient. When the time series have different change patterns or the correlation has a lag relationship, these methods cannot reflect the correlation well [1]. By extracting relevant features of time series and then calculating the correlation coefficient, but when the time series is very long, it may be difficult to give the fluctuation relationship of the time series accurately [2]. For time series of the same length, move to calculate the correlation coefficient to determine whether they have lag correlation, and find out the lag time. In addition to the correlation coefficient method, [4] Try to find out the influence structure diagram (SIG) when the time series is abnormal, and find the influence relationship between the time series. The time series anomaly is calculated by the skewness of the data in the sliding window relative to the distribution of the entire time series, but cannot recognize anomalies caused by some historical data. In addition, there are many statistical models used to calculate the relevance of time series [5–7], but they are aimed at specific time series and are not applicable.

Some researchers use Granger causality models for anomaly detection [10]. A graphical model to study Granger causality in multivariate time series is proposed to detect anomalies. The model calculates a “relevant anomaly” score for each variable through the Granger graphical model, which can provide information about the possible causes of anomalies. [11] proposed an improved Granger-Lasso algorithm, which is applicable to a wider range of heterogeneous time series [14]. The use of Granger causality test for alarm correlation is applicable to alarms based on anomalies, but there is no analysis method for the causes of anomalies.

In general, much work has been done on time series anomaly detection, but it still faces many challenges, including:

- (1) The influence relationship of time series is diverse, and it may be related in trend, shape, or pattern. At present, there is lack of more general models and methods to accurately determine the relevance of time series.
- (2) For anomaly detection based on telemetry data, due to the huge amount of data, the current real-time nature of many anomaly detection methods is difficult to meet the requirements.

This paper proposes a method of time series association analysis and anomaly detection based on Granger causality model. The main work is as follows:

- (1) Use the Granger causality model to discover whether there is a correlation between time series. This correlation is diverse, including trend correlation, time correlation, pattern correlation, and so on.
- (2) Improve the Granger causality model to reduce the time complexity of the calculation, so that it can quickly produce results for large amounts of data.
- (3) Quickly find out the cause of anomalies based on the Granger causality model.

The experimental analysis on the telemetry data of the space bridge shows that the method in this paper has a good performance on anomaly detection and anomaly detection.

2 Granger Causality Model

2.1 Correlation Analysis of Time Series

Clive W. J. Granger proposed the Granger causality model [13] in 1969 to analyze the correlation between economic variables. After years of development, Granger causality model has been applied to many other fields, and many improved methods have been proposed. The Granger causality model relies on the variance of the best least squares prediction.

Given two time series $\{X_s, i = 1, 2, 3, \dots, t\}$, $\{Y_s, i = 1, 2, 3, \dots, t\}$, t is the length of the time series, ε_{1s} is the white noise. p and q are the maximum lag periods of time series Y and X , respectively.

The test steps of Granger causality model are as follows:

- ① First verify whether X is the Granger cause of Y change. This test requires the estimation of the following regression model:

$$Y_s = \sum_{i=1}^p \alpha_i Y_{s-i} + \sum_{i=1}^q \beta_i X_{s-i} + \varepsilon_{1s} \tag{1}$$

$$Y_s = \sum_{i=1}^p \alpha_i Y_{s-i} + \varepsilon_{1s} \tag{2}$$

Equation (1) is called an unconstrained regression model (u), and Eq. (2) is called a constrained regression model (r).

- ② Propose the null hypothesis of the test: $H_0 : \beta_1 = \beta_2 = \dots = \beta_q = 0$, That is, X is not the Granger cause of the change of Y .
- ③ Calculate the sum of squared residuals of Eq. (1) and Eq. (2) RSS_u, RSS_r .
- ④ Construct F statistics, n is the sample size:

$$F = \frac{(RSS_r - RSS_u)/q}{RSS_u/(n - p - q - 1)} \sim F(q, n - p - q - 1) \tag{3}$$

- ⑤ Test the null hypothesis, if $F \geq F_\alpha$, then $\beta_1, \beta_2, \dots, \beta_q$ are not significantly zero, then the null hypothesis should be rejected, that is, X has an effect on Y ; otherwise, the null hypothesis is accepted, that is, X has no effect on Y .
- ⑥ Swap the positions of X and Y , and test whether Y is the Granger cause of the change of X in the same way.

From the above, we can see that if the variables are interrelated, for both tests, we should reject the null hypothesis, that is, under the condition that contains the past information of the variables X and Y , effect of prediction on Y is better than the prediction effect made by Y 's past information alone for Y , and the same is true for X . Therefore, we can determine whether there is a correlation between time series through a simple observation of the results.

Verifying the relationship between the two time series by using Granger Causality Model. For two time series, there may be four cases:

- (1) The first case is that X is the Granger cause of the change of Y , but Y is not the Granger cause of the change of X . We can say that X has a unilateral effect on Y .
- (2) The second case is that Y is the Granger cause of the change of X , but X is not the Granger cause of the change of Y . We can say that Y has a unilateral effect on X .
- (3) The third case is that X is the Granger cause of the change of Y and Y is also the Granger cause of the change of X . we can say that X and Y are related to each other.
- (4) The fourth case is that X is not the Granger cause of the change of Y and Y is also not the Granger cause of the change of X . We can say that X and Y are independent of each other, and there is no influence relationship.

If the past information of variables X and Y is included, the prediction effect of variable Y is better than the prediction effect of Y by past information of Y alone, that is, variable X helps to explain the future change of variable Y , It is considered that the

variable X has an influence on the variable Y . The Granger causal correlation model relies on the variance of the best least squares prediction using all the information at certain points in the past, whether X and Y are trend correlations, shape correlations, or pattern correlations. All can be verified by Granger causality test.

2.2 Anomaly Detection

If there is a correlation between time series, then when they are in a normal state, if conduct Granger causality test, the result of the test is the case (3), that is, the two variables affect each other. For real-time monitoring, these variables are monitored in real time through a sliding window. Once it is found that the time series are not related to each other, an abnormality can be detected.

The prerequisite for Granger causality test is that the time series must be stable, otherwise false regression problems may occur. Therefore, before the Granger causality test, we should use the unit root test should test the stationarity of the time series of each indicator. The ADF test is used to check whether there is a unit root in the sequence: if the sequence is stationary, there is no unit root; otherwise, there will be a unit root. Therefore, the null hypothesis of the ADF test is that there is a unit root. If the statistical significance of the test is less than the three confidence levels (10%, 5%, 1%), then there are 90%, 95%, and 99% to reject the null hypothesis. If the data is unstable, differential processing is necessary until the data is stable.

Then use the processed data to test Granger causality. Take a sliding window of length T to detect time series that have been determined to influence mutually. By moving the window, conduct Granger causality test on each pair of time series of length T . Once it is found that the result of the test is not the case (3). That is, if there is no relationship that affects each other, an abnormality occurs, and an alarm is immediately issued to further determine the cause of the abnormality.

2.3 Analysis of Exception Causes

When the Granger causality model test is carried out through the sliding window and an abnormality is found, the cause of the abnormality is determined by analyzing the test results. If the test result is the case (1), X has an impact on Y , but Y has no impact on X , we can judge that X is abnormal, because at this time X appears a fluctuation different from the normal state, and the state of Y is still the same as in the normal state; if the test result is the case (2), Y has an impact on X , but X has no impact on Y . We can judge that Y is abnormal, because at this time Y has fluctuated differently from the normal state, and the state of X is still the same as in the normal state. If it is the case (4), further discussion is needed, which is the content of future research.

3 Case Study

In order to verify the effectiveness of the method in this paper, two sets of related time series X_1, X_2 are taken. As shown in Fig. 1, it can be seen that the time series X_1 and X_2 are related.

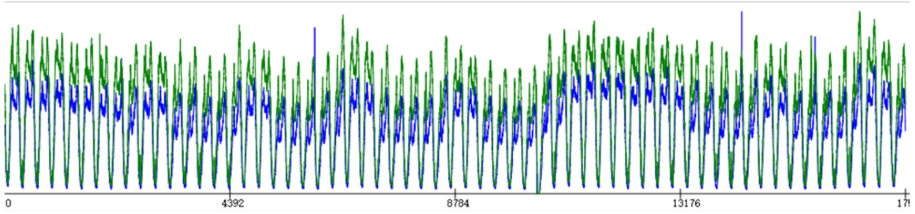


Fig. 1. Time series X_1, X_2

By conducting Granger causality test for X_1, X_2 . The results show that X_1 and X_2 have a mutually influential relationship, which is very consistent with the actual situation, which can explain that the Granger causality model is effective in judging the correlation of time series.

When X_1 or X_2 is abnormal, take a sequence of time, and use sliding window to detect anomalies. When we move the sliding window, the result of Granger causality test during the abnormal time is no longer the mutually influential relationship between X_1 and X_2 , the result matches the true situation. it can be concluded that even in the case of a large amount of data, the method in this paper can detect anomalies in time.

In order to verify whether the cause of the abnormality can be detected, firstly take the time series of X_1 is abnormal but X_2 is normal, and the results of Granger causality test are shown in Table 1. It can be concluded that the result of the test is that X_1 has an effect on X_2 , but X_2 has no effect on X_1 , and it is judged that X_1 is the cause of exception, which is in line with the actual situation.

Table 1. Test result 1

Lag	1	2	3	4
The null hypothesis: X_2 has no effect on X_1 (P value)	0.2051	0.6559	0.1987	0.2017
The null hypothesis: X_1 has no effect on X_2 (P value)	0.0052	0.0001	0.0000	0.0000

Then take the time series where X_2 is abnormal but X_1 is normal, and the result is shown in Table 2. It can be seen that the result of the test is that X_2 has an effect on X_1 , but X_1 has no effect on X_2 , and it we can judge that X_2 has an abnormality, which is in line with the actual situation. It follows from this that the method in this paper is effective in determining the cause of anomalies.

Table 2. Test result 2

Lag	1	2	3	4
The null hypothesis: X_2 has no effect on X_1 (P value)	0.0056	0.0005	0.0006	0.0005
The null hypothesis: X_1 has no effect on X_2 (P value)	0.3131	0.5377	0.3537	0.2956

4 Conclusion

This paper attempts to use the Granger causality model to determine whether the time series are correlated, based on the found relationship to monitor in real time, find anomalies in time, and further determine which variable caused the anomaly when the anomaly occurs. The case study shows that the proposed method in this paper is effective. In future work, we will further study how the related variables affect each other, as well as the analysis of the influence of one variable on other related variables when abnormal.

References

1. Su, Y., et al.: CoFlux: robustly correlating KPIs by fluctuations for service troubleshooting. In: Proceedings of the International Symposium on Quality of Service, IWQoS 2019 (2019)
2. Sakurai, Y., Papadimitriou, S., Faloutsos, C.: BRAID: stream mining through group lag correlations. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 599–610 (2005)
3. Cohen, J.: Statistical Power Analysis for the Behavioural Sciences (1988)
4. Oliner, A.J., Kulkarni, A.V., Aiken, A.: Using correlated surprise to infer shared influence. In: 2010 IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 191–200. IEEE (2010)
5. Hamao, Y., Masulis, R.W., Ng, V.: Correlations in price changes and volatility across international stock markets. *Rev. Financ. Stud.* **3**(2), 281–307 (1990)
6. Lütkepohl, H.: Forecasting with varma models. In: Handbook of Economic Forecasting, vol. 1, pp. 287–325 (2006)
7. Harris, R.I.D.: Using Cointegration Analysis in Econometric Modelling. Prentice Hall, Upper Saddle River (1995)
8. Fan, J., Yao, Q.: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer (2008)
9. Shanbhag, S., Wolf, T.: Accurate anomaly detection through parallelism. *IEEE Netw.* **23**(1), 22–28 (2009)
10. Qiu, H., Liu, Y., Surahmnya, N.A., et al.: Granger causality for time-series anomaly detection. In: IEEE 12th International Conference on Data Mining, pp. 1074–1079 (2012)
11. Behzadi, S., Hlaváčková-Schindler, K., Plant, C.: Dependency anomaly detection for heterogeneous time series: a granger-lasso approach. In: IEEE International Conference on Data Mining Workshops, ICDMW, November 2017, pp. 1090–1099 (2017)
12. Saha, H., Liu, C., Jiang, Z., Sarkar, S.: Exploring Granger causality in dynamical systems modeling and performance monitoring. In: Proceedings of the IEEE Conference on Decision and Control (CDC), December 2018, pp. 2537–2542 (2019)

13. Granger, C.W.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3), 424–438 (1969)
14. Qin, X., Lee, W.: Statistical causality analysis of INFOSEC alert data. In: Vigna, G., Kruegel, C., Jonsson, E. (eds.) RAID 2003. LNCS, vol. 2820, pp. 73–93. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45248-5_5