ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

Time series processes monitoring using statistical process control and machine learning approaches with applications on social network surveillance*

Abdaljbbar B. A. Dawod[1b]

[b]Faculty of Informatics, University of Debercen, Hungary.
dawod.abdaljbbar@inf.unideb.hu

Abstract

The surveillance of time series processes aims to detect any unexpected drifts/shifts in the process compared to a predefined standard settings. Consequently, statistical process control (SPC) is a statistical approach that was invented for process monitoring purposes, and the control charts are considered as the most powerful tool for SPC. The basic structures of control charts usually assume that, the process generates independent and normally distributed observations. Due to the autocorrelation characteristic of the time series, the original structures of control charts is inadequate to monitor time series process as it does not fulfill the independence assumption. Hence, monitoring the time series process without removing that autocorrelation will be misleading and gives false alerts of shift detection in the process [3].

One of the prominent methods used to modify the original structures of the control charts that will suit time series process is the model-based approach. The method basically uses an adequate time series model that fits the process, then utilize its residuals as a monitoring process. Therefore, model-based approach is very sensitive to the adopted time series model, and it is important to pick a suitable model that fits the process and that can be used for different designs of control

*This research was supported by the doctoral school of informatics, university of Debercen
[1]PhD student, under the supervision of professor György Terdik

charts. Hence, this study intends to study the performance of the three popular types of control charts namely: Shewhart, exponentially weighted moving average, and cumulative sum. Some time series models will covered that include autoregressive (AR), moving average (MA), and autoregressive-moving average (ARMA) as potential models to fit the process of interest. Beside that, classification algorithm will be used as a supervised machine learning approach for detecting out-of-control processes [1, 2, 4].

The research expects to generate control charts that adequately will be used to monitor time series processes, and the performance of the new control charts will be investigated numerically using Monte Carlo simulations. Two performance aspects will be used as performance measures namely efficiency and robustness. Average run length (ARL) is used as an efficiency measure for different in-control and out-of-control states of the time series processes with different levels of autocorrelation.

Finally, the simulated results will be validated with a real application of monitoring the monthly active users of facebook social network (Source of the data: Facebook quarterly number of MAU (monthly active users) worldwide 2008-2022 Published by S. Dixon, Oct 27, 2022, https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/, last access of the website: 9^{th},Jan 2023).

References

[1] C. Bilen, X. Chen: *Comparison of control charts for autocorrelated process control*, International Journal of Quality Engineering and Technology 1.2 (2009), pp. 136–157, doi: 10.1504/ijqet.2009.031127.

[2] A. B. Dawod, M. Riaz, S. A. Abbasi: *On model selection for autocorrelated processes in statistical process control*, Quality and Reliability Engineering International 33.4 (2017), pp. 867–882, doi: https://doi.org/10.1002/qre.2063.

[3] C.-W. Lu, M. R. Reynolds Jr: *Control charts for monitoring the mean and variance of autocorrelated processes*, Journal of Quality Technology 31.3 (1999), pp. 259–274, doi: 10.1080/00224065.1999.11979925.

[4] S.-V. Oprea, A. Bâra: *Machine learning classification algorithms and anomaly detection in conventional meters and Tunisian electricity consumption large datasets*, Computers & Electrical Engineering 94 (2021), p. 107329, doi: 10.1016/j.compeleceng.2021.107329.