

Practical Considerations for Implementing the Isolation Forest EWMA Control Chart in Phase II Process Monitoring

T. Perdakis¹, G. Celano², S. Chakraborti³

¹Department of Statistics, Athens University of Economics and Business Athens, Greece

²Department of Civil Engineering and Architecture, University of Catania, Italy

³Department of Information Systems, Statistics and Management Science, University of Alabama, Tuscaloosa, AL, USA

Statistical monitoring methodologies are constantly evolving to cope with various data complexities, non-standard situations and related questions that often arise in modern day applications. These questions, for example, relate to the speed of data collection, multi-dimensionality and distributional assumptions. In this context, researchers in the Statistical Process Monitoring (SPM) literature have recently considered Statistical Learning (SL) techniques as viable means to define models under more practical assumptions, and thus to allow setting up control charts for monitoring process stability in a variety of situations. However, a rigorous investigation of some of the key issues related to the implementation of the SL based control charts, supporting on-line (Phase II) SPM, has been lacking and is much needed. As a first step in this direction, here we consider a control chart based on the Isolation Forest (IF), an unsupervised SL technique running an ensemble of decision trees, which has recently been extended to the SPM area. We examine key implementation issues related to the selection of a proper Phase I reference sample, as it strongly influences the model trained to construct the control chart and its statistical performance properties. Our results show that correctly running an IF control chart is a challenging task, needing careful attention by the practitioners. In particular, a huge Phase I sample size and a careful check of the reference sample stability are required to maintain the in-control performance of the control chart at the anticipated target level, and to prevent a significant deterioration of the chart's shift detection capability.

Keywords: Statistical Learning, Multivariate processes, Isolation Forest, Statistical Performance

Acknowledgements

The work of Prof. G.Celano in this research was funded by Università degli Studi di Catania under the Grant Scheme PIA.CE.RI. Linea 1 2024-2026, project ISTRICE, ID: 59722022316.