Alternatives to Statistical Process Control for Automatic Alerts Generation

Introduction

There is evidence that Statistical Process Control (SPC) based approaches to alert creation can produce early alerts to systematic change such as influenza outbreaks [1] and anomaly detection in patient records [2]. While not as powerful as more sophisticated methods such as Convolutional neural networks [4], Graph Neural Networks [5] or even Markov Chaining [6], they do provide a simple manual rule-based approach to generating alerts that should be superior to Red, Amber, Green (RAG) ratings when there is a limited amount of data available. Though much like RAG ratings, an assumption of normality must be made about the data of interest.

There are many reasons a predictive model, including SPCs, may produce false outcomes on historical, new or future scenarios but they largely fall into 3 categories:

1. The model has been insufficiently constructed. This could be due there being insufficient data or an insufficient understanding of the processes that produce alerts. In simpler terms, the model is under-trained.

2. The rules involved in generating an output attempt to capture every possible feature of the data, even extremely niche or outlying scenarios. In simpler terms, the model is over-trained.

3. The data is unsuitable (e.g., Dynamic, sparse or with underlying systematic behaviour). In simpler terms, the model is untrainable.

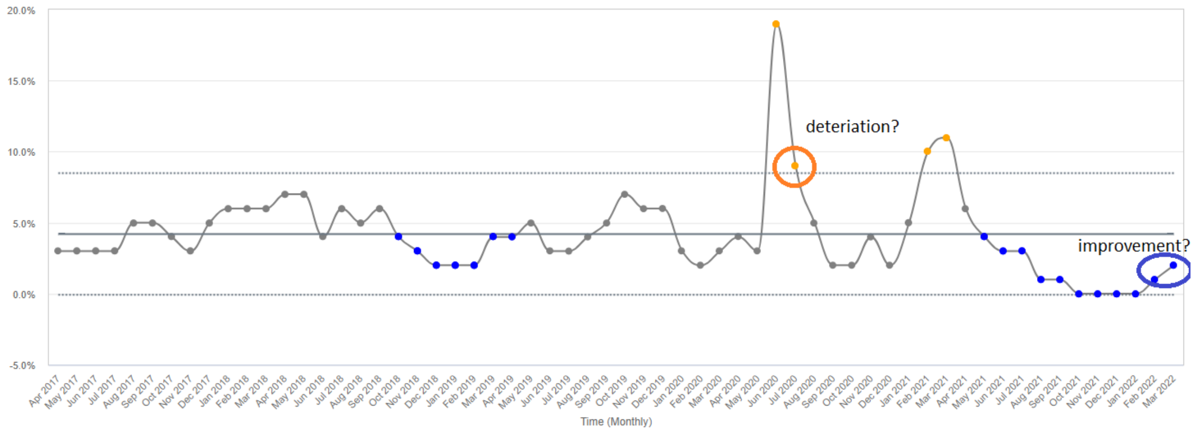
In complete real-world scenarios, with real-world data, some false categorisation of datapoints is inevitable. The main task is to use the available data and the available understanding of the processes to produce a predictive model that is as accurate as possible with current data but is also sufficiently general for future data. SPCs are an example of a modelling tool that produces alerts to exceptional values based on previous values.

Current Alert Types and their Issues

SPC charts and alerts are currently being generated in many industrial scenarios such as healthcare [7] and manufacturing [8]. The main feature of SPCs is that they take a mechanical, objective, unsupervised approach. They mechanise the visual process whereby we look at plots to decide if there is anything ‘out of the ordinary’.

False Positives

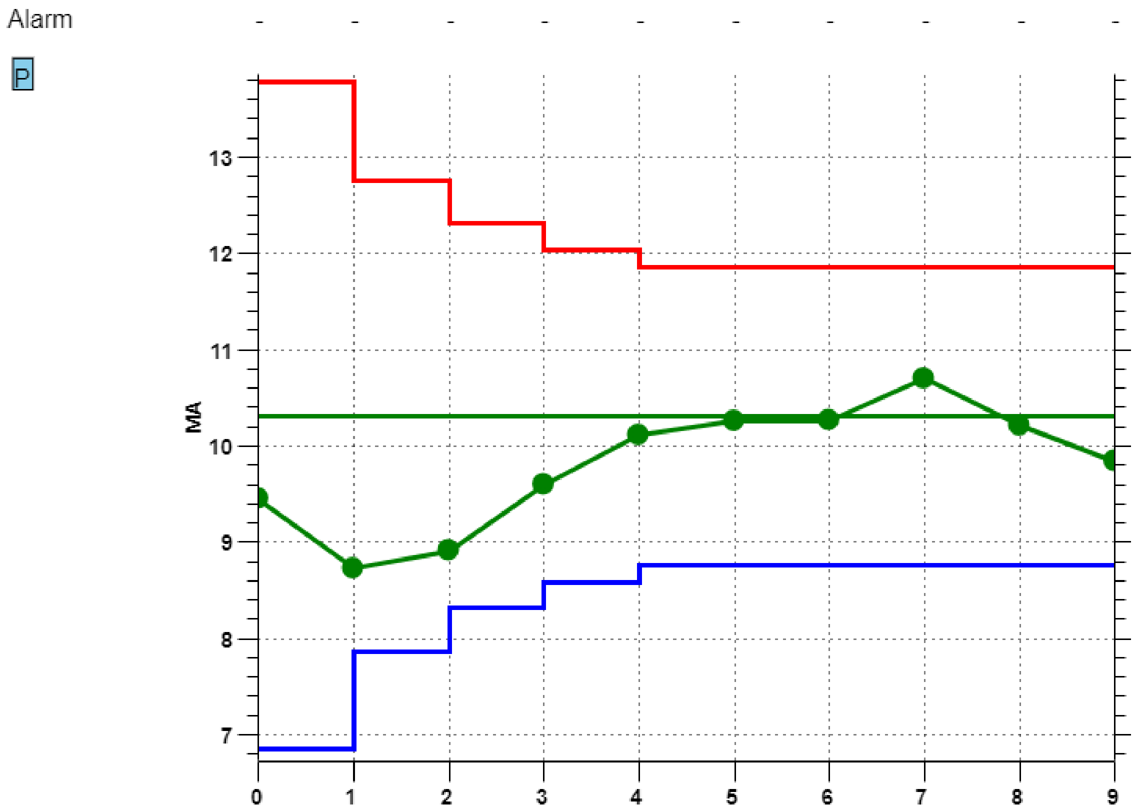
Using a set of mechanistic rules that operate on a small number of datapoints in isolation is simple and transparent, but we should accept that such a hardcoded, trivial approach to quantifying a qualitative measurement will yield some false positives (alert when a user thinks one wasn’t justified) and false negatives (NO alert when a user thinks one WAS justified).



When there is currently no ‘Ground Truth’ about nature of a correct alert, only unsupervised approaches such as SPCs can be used.

The normality assumption

There is an underlying agreement that SPCs are best used on normally distributed data (e.g., (“Conventional SPC charts require the normality assumption on the process response distribution. In reality, this assumption is often invalid.” [3]). Using SPCs on moving average (MA) data is possible, though it challenges the normality assumption and also means different confidence limits should be used. E.g.,



Unless the data has some underlying periodic or self-similar processes, calculating a moving average will reduce the variance in proportion to the window size [9]. This needs to be accounted for in the confidence limits.

SPCs are poor for non-linear outcomes

In addition to the previous example where there changes in local circumstances create a step change (that in turn creates alerts), the following example also shows how non-linear outcomes are difficult to interpret in SPC. In general, where changes in a dependent variable have a complex meaning, rag ratings and SPCs need to be carefully applied, or not applied at all.

Sub-Optimal Control Limits can be generated

Sometimes the formula for producing control limits produces less than ideal values, the following has an upper limit of >100% and a lower limit of <0%, neither is possible. Though alerts could still be generated using other rules, it would seem more sensible to generate control limits that are within the realms of possibility.

Machine learning solutions

This project would take current SPC theory as starting point and investigate new approaches to automated prediction using more intelligent approaches that take an adaptive (learning from previous values in the timeline) and transfer learnt (learning for similar scenarios) approach. Quantitative experiments will be carried out to understand the effectiveness of novel approaches when compared to different forms of SPCs

Supervisory Team

Dr Prapa Rattadilok

Professor Farshid Amirabdollahian

Contact

For informal enquiries about this PhD, please contact Dr Prapa Rattadilok p.rattadilok@herts.ac.uk

References

[1] Gren, Lisa H., et al. "Point-of-care testing as an influenza surveillance tool: methodology and lessons learned from implementation." Influenza Research and Treatment 2013 (2013).

[2] Kassakian, Steven Z., et al. "Clinical decisions support malfunctions in a commercial electronic health record." Applied clinical informatics 8.03 (2017): 910-923.

[3] Qiu, Peihua, and Jingnan Zhang. "On phase II SPC in cases when normality is invalid." Quality and Reliability Engineering International 31.1 (2015): 27-35.

[4] Chirra, Venkata Rami Reddy, Srinivasulu Reddy Uyyala, and Venkata Krishna Kishore Kolli. "Deep CNN: A Machine Learning Approach for Driver Drowsiness Detection Based on Eye State." Rev. d'Intelligence Artif. 33.6 (2019): 461-466.

[5] Liu, Peng, et al. "Sensor network prediction based on spatial and temporal GNN." ITM Web of Conferences. Vol. 47. EDP Sciences, 2022.

[6] Roshan, Gholamreza, and Panagiotis T. Nastos. "Assessment of extreme heat stress probabilities in Iran's urban settlements, using first order Markov chain model." Sustainable cities and society 36 (2018): 302-310.

[7] Gray, William K., et al. "Identifying unwarranted variation in clinical practice between healthcare providers in England: Analysis of administrative data over time for the Getting It Right First Time programme." Journal of Evaluation in Clinical Practice 27.4 (2021): 743-750.

[8] Madanhire, Ignatio, and Charles Mbohwa. "Application of statistical process control (SPC) in manufacturing industry in a developing country." Procedia Cirp 40 (2016): 580-583.

[9] Roadknight, Chris, Ian Marshall, and George Bilchev. "Network performance implications of variability in data traffic." BT technology journal 18.2 (2000): 151-158.