Time Series Anomaly Detection in Healthcare using Variogram Transition Rules

Janne Cools [1], Toon Boeckling [1], and Antoon Bronselaer [1]

[1] Dept. of Telecommunications and Information Processing, University of Ghent, Belgium
{janne.cools,toon.boeckling,antoon.bronselaer}@ugent.be

In healthcare, several physiological parameters of patients, such as body weight, temperature and blood pressure, are measured and recorded. Since each measurement is timestamped, they can be represented as time series data. An anomaly in such data is a point that deviates from the expected behavior. Healthcare data may contain these errors due to manual input or device malfunction. Over the past decades, identifying anomalies has become an increasingly important research topic, leading to significant advancements in anomaly detection algorithms [1]. Many of these algorithms assume that the time series under consideration is regular and dense, meaning that measurements occur at constant and short time intervals. When the patient parameters are monitored by machines, this assumption holds. However, EPD (Electronic Patient Dossier) data are manually updated by nurses who record parameters themselves. This data is often sparse and irregular, as measurement times vary depending on the patient's condition and the workload. Moreover, manual entry increases the likelihood of typographical mistakes or estimated values.

The focus of this research is on detecting and correcting anomalies in EPD data, which features irregularity and sparseness. Our approach is based on the selection rule framework [2, 3, 4]. In this framework, selection rules represent combinations of values that are not permitted to appear in a data object, which, in the context of EPD time series, records measurements taken at a specific time point. To detect anomalies, however, measurements taken over multiple time points need to be compared. To address this, we propose to extend selection rules to *transition rules*, which define a maximum allowed change in value for a parameter between two successive data points. Because the data are irregularly spaced, applying a single transition rule would overgeneralize the allowed value difference, which depends on the elapsed time. Hence, a parameter is likely to be covered by several transition rules, where each transition rule specifies a value difference based on a time difference. For example, one rule defines an allowed weight difference of 3 kilograms within a day, while another defines a difference of 5 kilograms within a week.

Since each parameter may have multiple transition rules, this increases the execution time, as the complexity of anomaly detection using selection rules directly depends on the number of rules [3]. To reduce the time complexity, we propose an extension of transition rules, called a *variogram transition rule*, which computes the allowed value difference as a function of the time interval. This is inspired by the variogram concept from geostatistics [5], which quantifies spatial variability, and offers several benefits. First, a large selection of variogram models, such as exponential, linear and Gaussian, is available for the computation. Next, combining multiple transition rules into one variogram transition rule decreases the execution time. Finally, all concepts of the selection rule framework can still be applied to this extension.

References

[1] Schmidl, S., Wenig, P., and Papenbrock, T. (2022). Anomaly detection in time series: a comprehensive evaluation. *Proceedings of the VLDB Endowment*, 15(9), 1779–1797.

[2] Boeckling, T., De Tré, G., and Bronselaer, A. (2022). Cleaning data with selection rules. *IEEE Access*, 10, 125212–125229.

[3] Boeckling, T., De Tré, G., and Bronselaer, A. (2022). Efficient edit rule implication for nominal and ordinal data. *Information Sciences*, 590, 179–197.

[4] Bronselaer, A., and Acosta, M. (2023). Parker: Data fusion through consistent repairs using edit rules under partial keys. *Information Fusion*, 100, 101942.

[5] Gringarten, E., and Deutch, C. V. (2001). Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, 33, 507–534.