Anomaly Detection For Monitoring

IKHTIAR AHMED
230890

Overview

- Anomaly Detection
- Types of Anomalies
- Time Series Data
- Modeling
- Prediction
- Charateristics
- Disavantages
- Conclusion

Anomaly Detection?

- To find the noisy metrics
- To identify the unexpected event

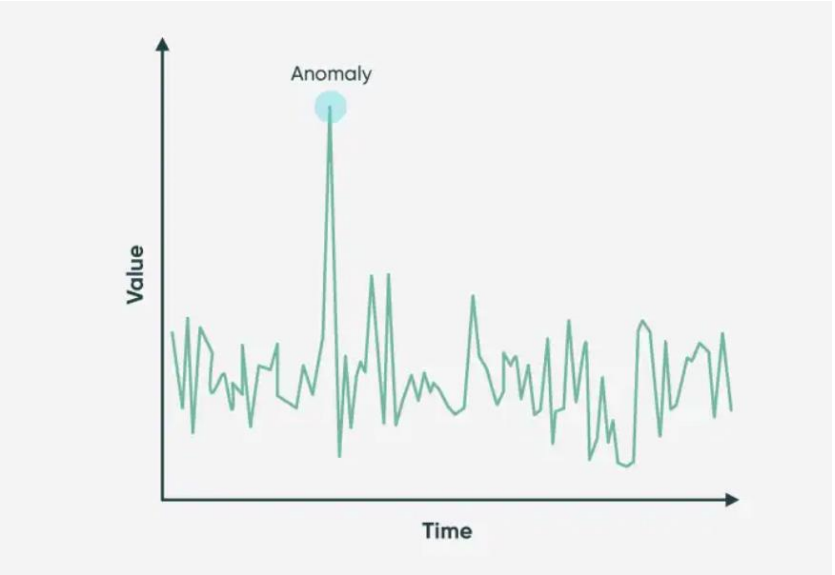

Fig: Anomaly Detection

Source : https://www.pyimagesearch.com/2020/01/20/intro-to-anomaly-detection-with-opencv-computer-vision-and-scikit-learn/

Examples

- Intrusion Detection
- Credit Card Fraud
- Health Monitoring

Types of Anomalies

- **Point Anomalies**
- **Contextual Anomalies**
- **Collective Anomalies**

Point Anomaly

- Single Instance
- Used to detect credit card Fraud

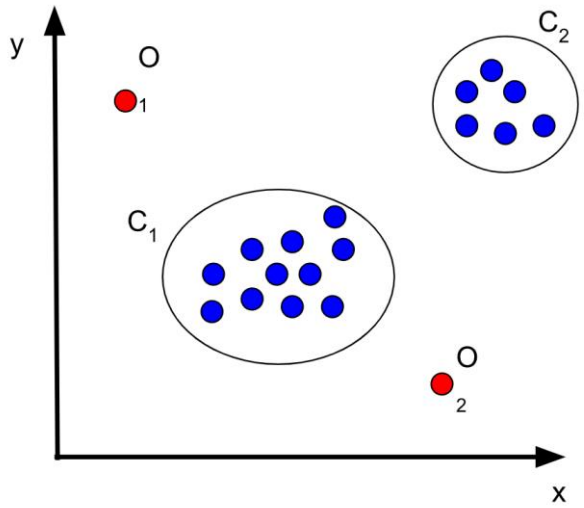

Fig: Point Anomaly

Source: https://elf11.github.io/2018/09/20/data-science-anomaly-detection.html

Contextual Anomaly

- An individual data instance is anomalous within a context
- Requires a notion of context
- Also referred to as conditional anomalies

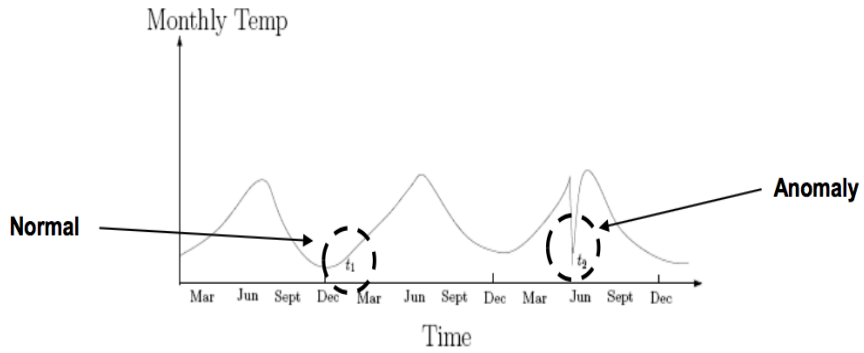

Fig: Contextual Anomaly

Source: https://www.repetico.com/card-67787341

Collective Anomaly

- A collection of instances
- Copy data from machine

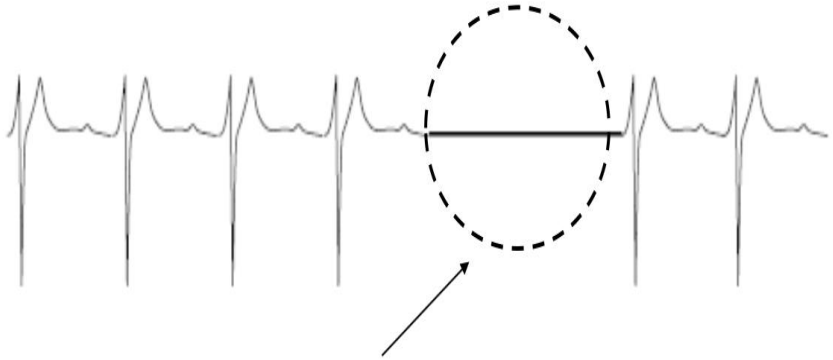

Fig: Collective Anomaly

Source: https://www.repetico.com/card-67787341

Time Series Data

- Sequence of data points
  - Univariate data
  - Multi-variate data

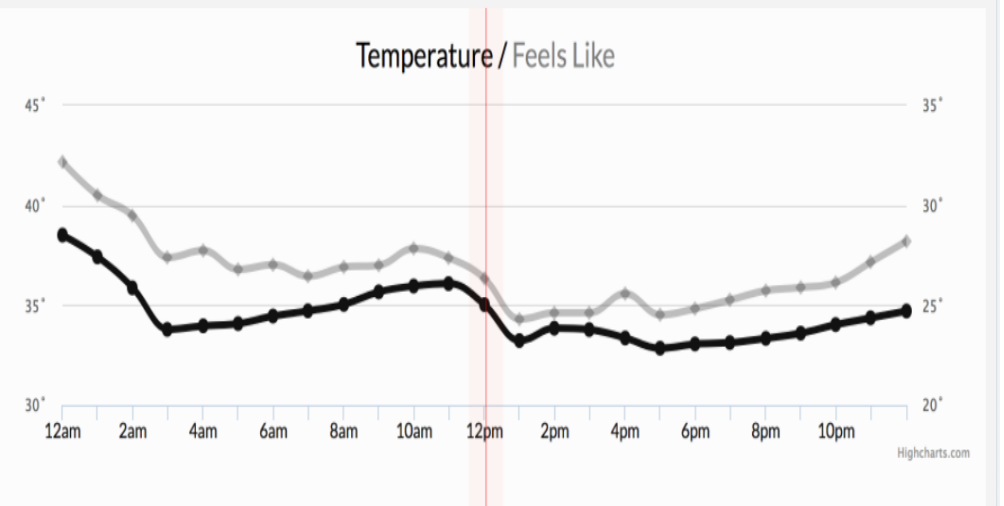

Fig: Time Series Data

Source: https://www.influxdata.com/what-is-time-series-data/

Stationarity and Differentiation

- **Stationarity**
  - **Properties do not depend on the time**

- **Non-Stationary Stochastic Process**
  - **Trend**
  - **Volatility**
  - **Seasonality**

Modeling

Statistical Process Control (SPC)

- Fixed Control Chart
- Moving window control chart
- Exponentially Weighted Moving Average (EWMA)

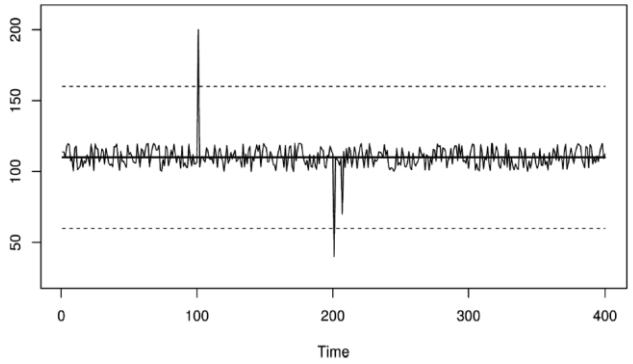

Fig: Basic control chart with fixed limit

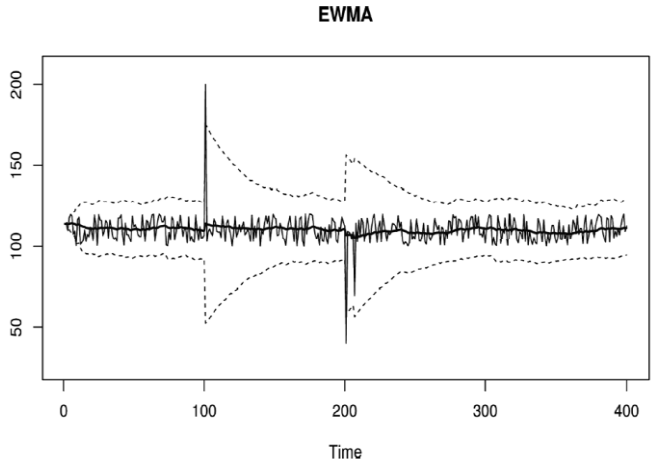

Fig: EWMA

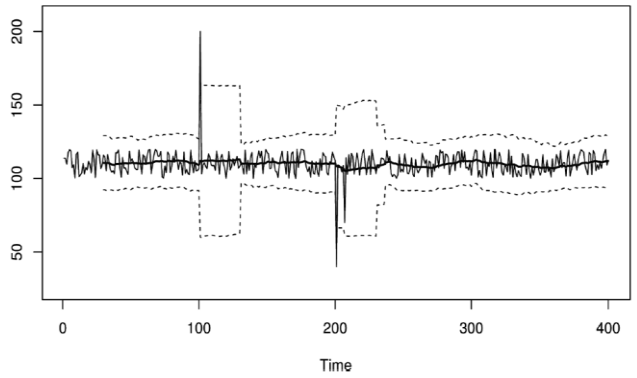

Fig: Moving Window control chart

Source: Baron Schwartz and Preetam Jinka “Anomaly Detection for Monitoring”

Modeling

- **Autoregressive Integrated Moving Average (ARIMA)**
  - Predict future values
  - Parameters (p,q,d)
  - Linear Regression

Prediction

- What it should be?
- Why do we need Prediction?
- How to get Prediction?

Characteristics

- **Trend**
  - **Historical changes**
  - **Continuously increasing or decreasing**
  - **Linear trend**

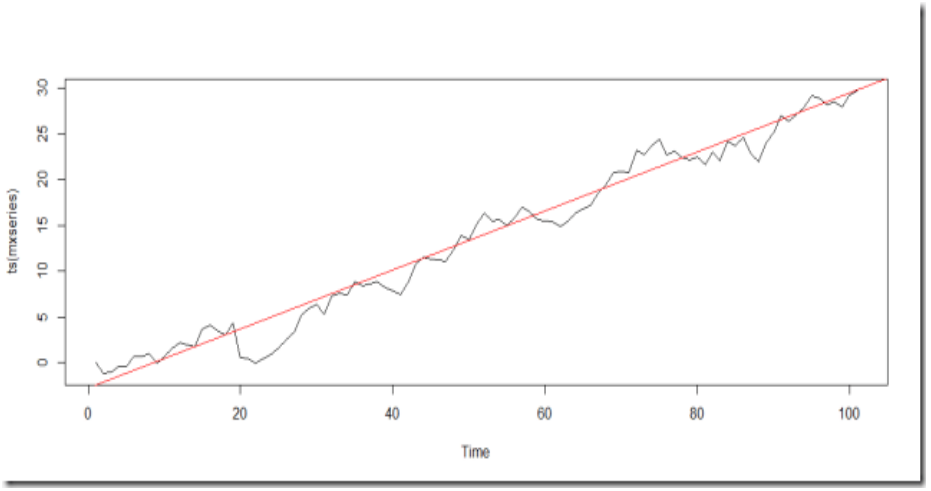

Fig: Time Series data with Linear Trend

Source: https://www.oraylis.de/blog/2015/trend-in-times-series-analysis

Characteristics

- **Seasonality**
  - **Regular and predictable changes such as weekly and monthly.**

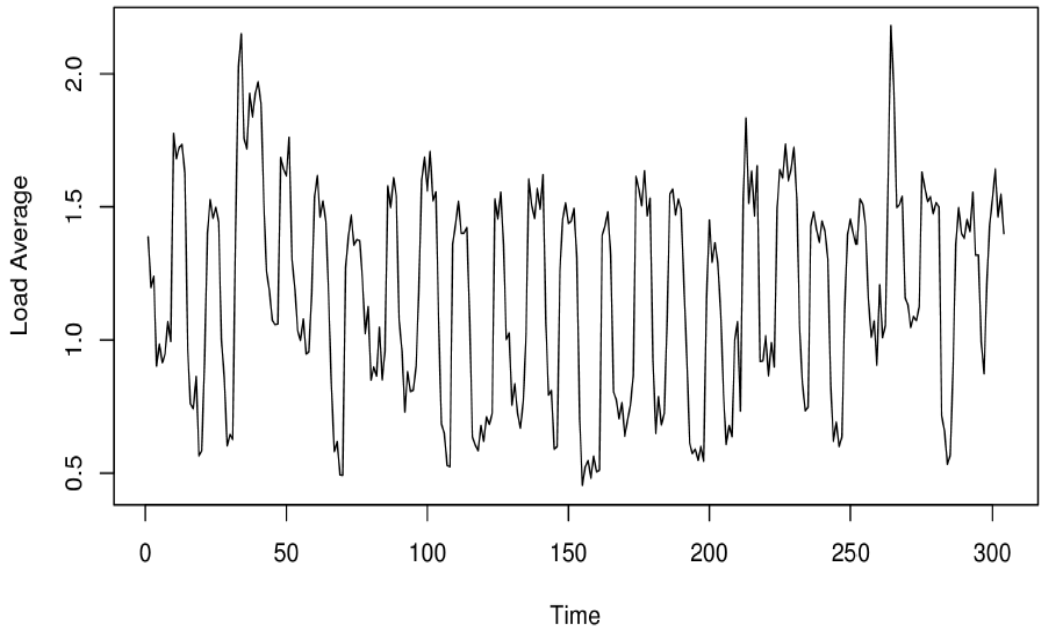

Fig: Dealing with Seasonality

Source: Baron Schwartz and Preetam Jinka "Anomaly Detection for Monitoring"

Why do we need Decomposition?

- **Additive Decomposition**
  - Constant over the time

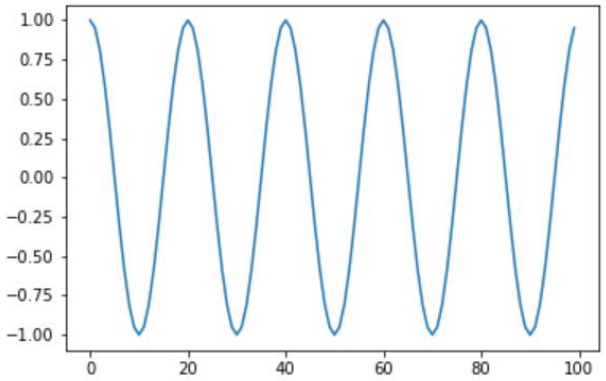

Fig: Additive Seasonality

- **Multiplicative Decomposition**
  - Increases over the time

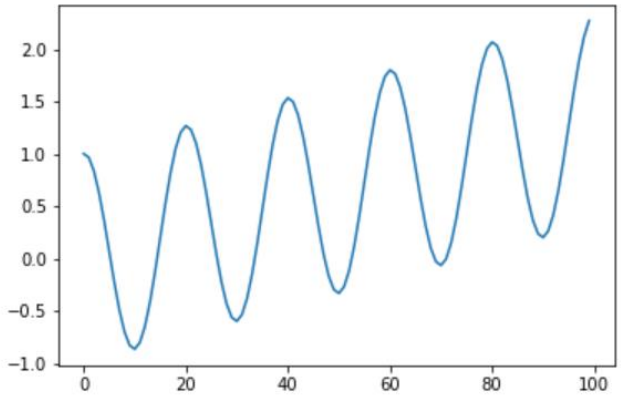

Fig: Multiplicative Seasonality

Source: https://towardsdatascience.com/finding-seasonal-trends-in-time-series-data-with-python-ce10c37aa861

Extract Pattern

- **Fourier Transform**
  - trigonometric functions sine and cosine

The Fourier transform is used in many domains –

- Sound processing
- Filtering Data
- System Identification

Metric

- It's important to be sure that the problem you're trying to detect has a reliable signal.

Example metrics we could check include-

- Error Rate
- Throughput
- Latency

Mean Shift Analysis

- Represents the fundamental changes to model parameter

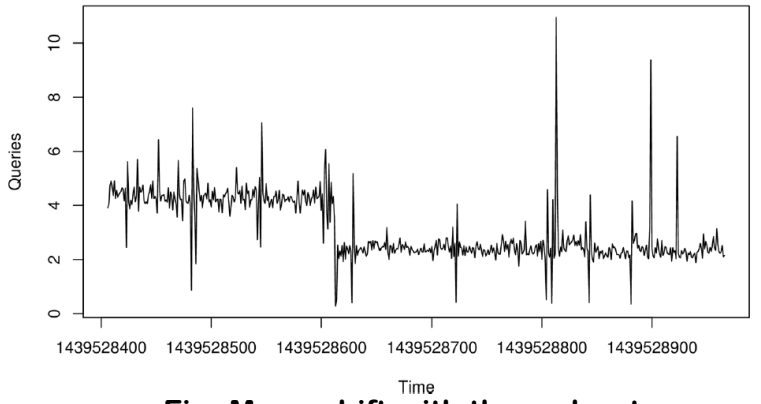

Fig: Mean shift with throughput

- Cumulative Sum Control Chart (CUSUM)
  - To detect small shifts from the process target

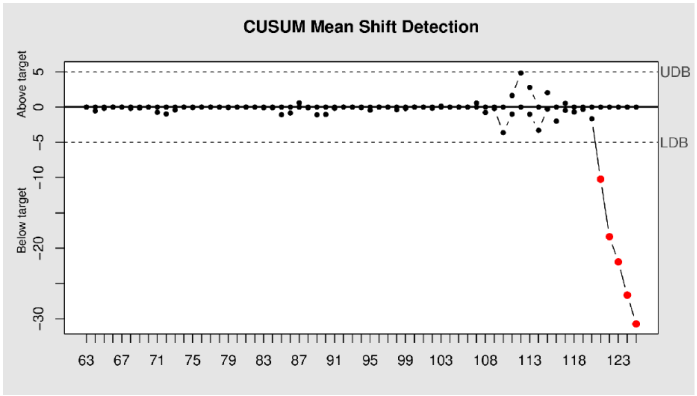

Fig: Cumulative Sum Control Chart

Source: Baron Schwartz and Preetam Jinka “Anomaly Detection for Monitoring”

Disadvantages

- Univariate data
- Common correlation over the time
- Autocorrelations and systematic time series

Research Outcome

- Common cause variation
- Special cause variation
- ARIMA Model
- Multivariate data
- Engineering process control

Conclusion

- Anomalies are common occurrences
- Different anomalies, various detection method
- Combining several methods
- It´s complicated

References

[1]. Baron Schwartz and Preetam Jinka “Anomaly Detection for Monitoring”

[2]. Layth C. Alwan and Harry V. Roberts “ Time-Series Modeling for Statistical Process Control”, January 1988, Vol. 6, No. 1

[3]. Dominique T. Shipmon, Jason M. Gurevitch, Paolo M. Piselli, Steve Edwards “ Detection of Anomalous Drops with Limited Features and Sparse Examples in Noisy Highly Periodic Data”

[4]. Anomaly detection “https://blogs.oracle.com/ai-and-datascience/post/introduction-to-anomaly-detection”

[5]. Time series data “https://www.investopedia.com/terms/t/timeseries.asp”

[6]. Seasonality Analysis and Forecast in Time Series “https://medium.com/swlh/seasonality-analysis-and-forecast-in-time-series-b8fbba820327”

[7]. Finding Seasonal Trends in Time-Series Data “https://towardsdatascience.com/finding-seasonal-trends-in-time-series-data-with-python-ce10c37aa861 ”

[8]. Lee-Ing Tong, Chien-Hui Yang, Cheng-Yi Huang, Cheng-Chi Shou “Integrating SPC and EPC for Multivariate Autocorrelated Process ”

References

[9]. Statistical process control and interrupted time series: a golden opportunity for impact evaluation in quality improvement “https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4680165/ ”

[10]. Autoregressive Integrated Moving Average “ https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp”

[11]. Stationarity and differencing “https://otexts.com/fpp2/stationarity.html ”

[12]. Deconstructing Time Series using Fourier Transform “ https://medium.com/@khairulomar/deconstructing-time-series-using-fourier-transform-e52dd535a44e”

[13]. Time series features extraction using Fourier and Wavelet transforms on ECG data https://blog.octo.com/time-series-features-extraction-using-fourier-and-wavelet-transforms-on-ecg-data/

[14]. Identification of Seasonality in Time Series with Python Machine Learning Client for SAP HANA “ https://blogs.sap.com/2020/12/18/identification-of-seasonality-in-time-series-with-python-machine-learning-client-for-sap-hana/”

[15] Richard Stone and Mark Taylor “Time Series Models in Statistical Process Control: Considerations of Applicability ”, 1995, Vol.44, No. 2 (1995), pp. 227-234