A review on outlier/anomaly detection in (univariate) time series data

Bruno Mendes - 02/03/2021

Outlier/Anomaly

- The analysis of outliers in time series data examines anomalous behaviors across time.
- They are observations that **do not follow the expected behavior**.

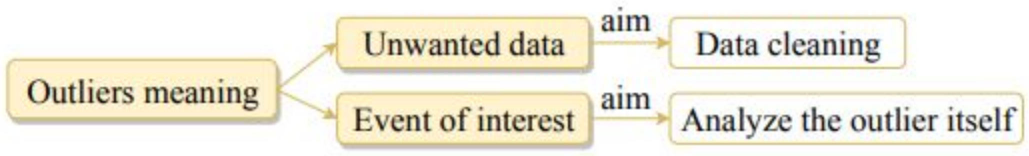


Fig. 1. Meaning of the outliers in time series data.

Outlier Type

- **Point outliers**: a point that behaves unusually in a specific time instant when compared either to other values in the time series, *global outlier*, or to its neighboring points, *local outlier*.

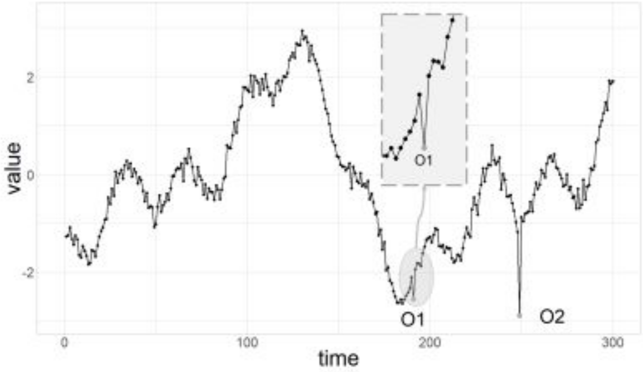


Fig. 2. Point outlier in an univariate time series.

Outlier Type

- **Subsequence outliers**: consecutive points in time whose joint behavior is unusual. The can also be *global* or *local*.

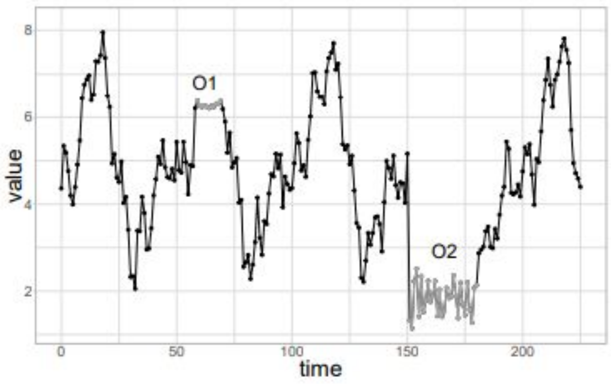


Fig. 3. Subsequence outlier in an univariate time series.

Point Outlier Detection

- The **most common** outlier detection task in time series data.
- Key characteristics:
  - **Temporality**: some methods consider the inherent temporal order of the observations. Those that do not produce the same results even if applied to a shuffled version of the series.
  - **Streaming**: some methods are able to detect whether or not a new incoming datum is an outlier as soon as it arrives. It only takes past data into consideration.

Point Outlier Detection in Univariate Time Series

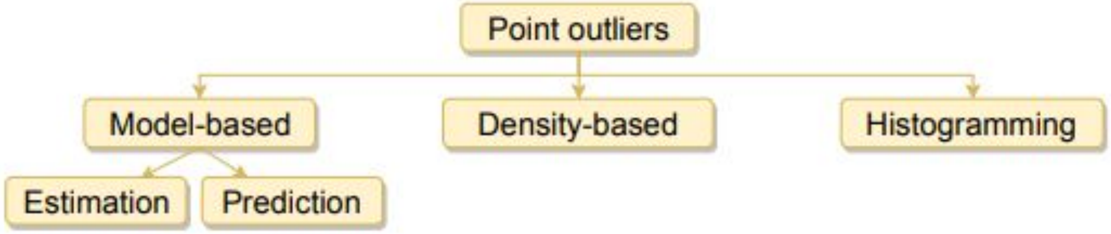


Fig. 4. Types of methods for detecting point outliers in univariate time series.

Model-based methods

- Defining **point outlier** as a point that significantly deviates from its expected value...
  - a point of time *t* can be declared an outlier if the distance to its expected value is higher than a predefined threshold □.

$$|x_t - \hat{x}_t| > \tau$$

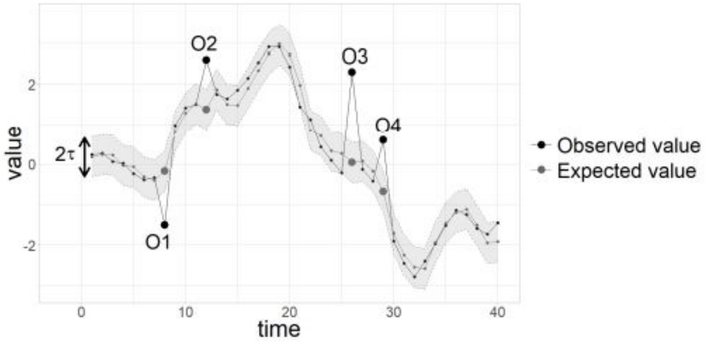


Fig. 5. Comparison between expected and observed values.

- They are all based on fitting a model.

Model-based methods

- **Estimation model-based**: if a prediction of a point is obtained by using both previous and subsequent observations.
- **Prediction model-based**: if it only considers past observations to make a prediction.

	Data used	$\longrightarrow$	Expected value	$\longrightarrow$	Point outliers
Estimation models	$\{x_{t-k_1}, \ldots, x_t, \ldots, x_{t+k_2}\}$	$\longrightarrow$	$\hat{x}_t$	$\longrightarrow$	$\lvert x_t - \hat{x}_t \rvert > \tau$
Prediction models	$\{x_{t-k}, \ldots, x_{t-1}\}$	$\longrightarrow$	$\hat{x}_t$		

Table 1. Model-based methods

Model-based methods

- Simple methods are based on constant models, where *basic statistics* are used to obtained the predicted value.
- Others intend to identify data points that are unlikely if a certain *fitted model* is assumed to have generated the data.
- Within the *prediction-based methods*, some may use a fixed model and thus are not able to adapt to the changes while others adapt to the evolution by *retraining the model*.

Density-based methods

- Defining **point outlier** as a point that is isolated from the majority...
    - a point of time *t* can be declared an outlier if it has less than □ neighbors, that is, when less than □ objects lie within *R* distance *t*.

$$x_t \text{ is an outlier} \iff |\{x \in X | d(x, x_t) \leq R\}| < \tau$$

- *d* is most commonly the Euclidean distance and *X* is the set of data points.

Density-based methods

- The detection of density-based outliers has been widely handled in non-temporal data **but** the concept of neighborhood is more complex in time series due to temporality.
    - When using a sliding windows, a point can be an outlier for a window but not for another.

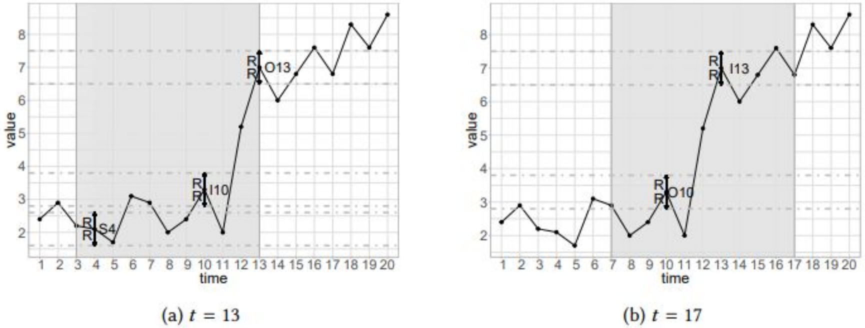
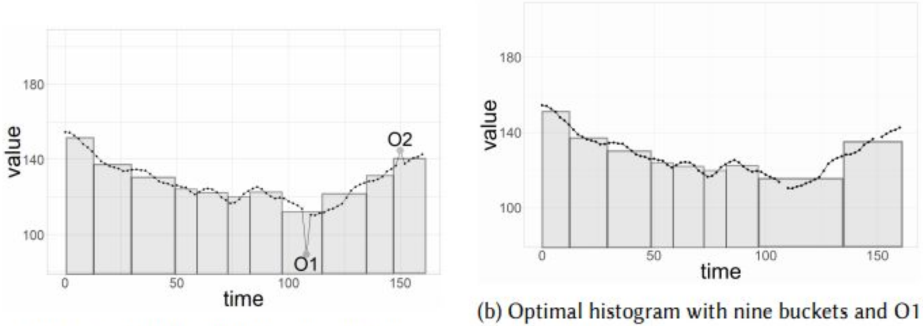


Fig. 6. Density-based outliers within a sliding window at time step t.

Histogramming methods

- It is based on detecting the points whose removal from the univariate time series results in a histogram representation with lower error than the original.



(a) Optimal histogram with eleven buckets.

(b) Optimal histogram with nine buckets and O1 and O2 removed.

Fig. 7. Example of a deviant set *D* = {*O*1, *O*2} in a univariate time series.

Subsequence Outlier Detection

- Key characteristics related to subsequence outliers:
  - They have a certain **length**, thus methods can consider fix-length subsequences or allow the detection of subsequences of different lengths. The number of subsequences will be higher when the length is shorter.
  - The data **representation** needs to be considered as comparison between subsequences is more challenging and costly than between points. Therefore, a representation of the subsequence is usually preferred over the original raw values.
  - They can be **periodic** if they repeat themselves over time.
- Subsequences consider the temporality by default.

Subsequence Outlier Detection in Univariate Time Series

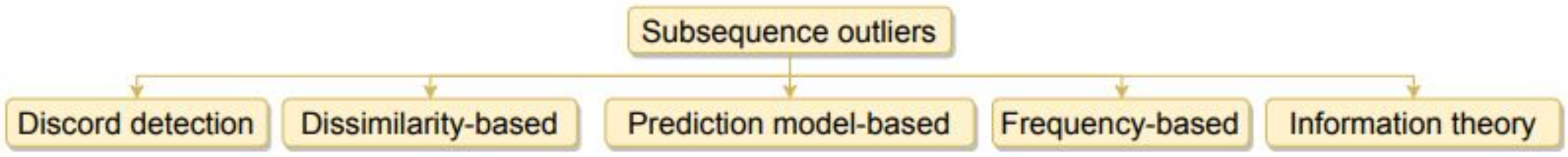


Fig. 8. Types of methods for detecting subsequence outliers in univariate time series.

Discord Detection methods

- Consists of detecting the most unusual subsequences in a time series **by comparing each subsequence with the others**.
- It typically requires the user to specify the length of the discord.

- The *simplest way* is to use brute-force.

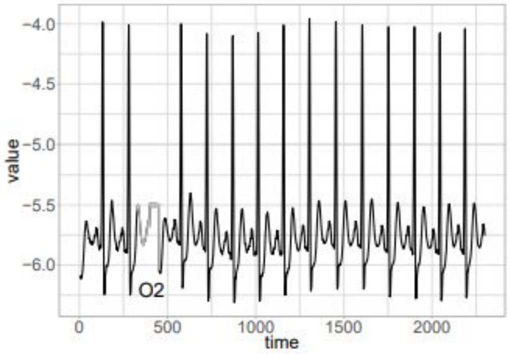
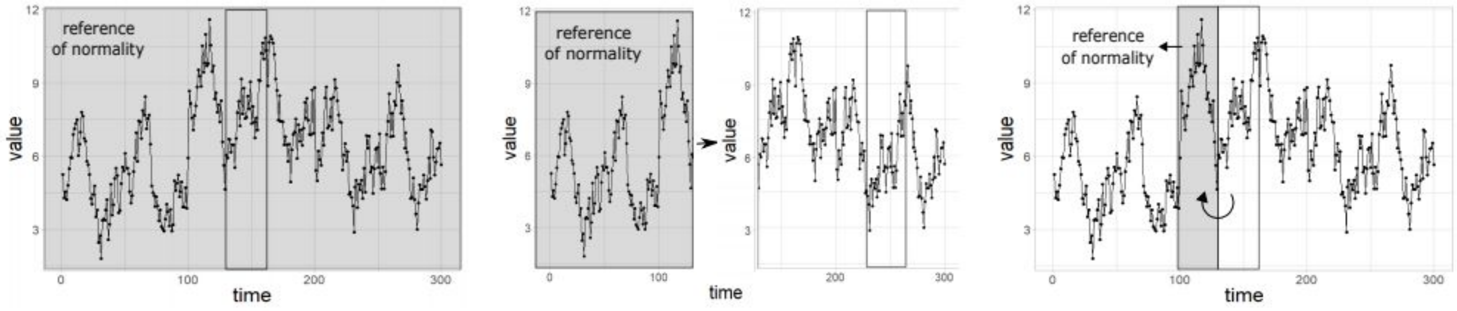


Fig. 9. Discord example.

Dissimilarity-based methods

- These methods are based on the **direct comparison of subsequences using a reference of normality**.
    - The reference of normality can vary widely, according with each case.



(a) Same time series. (b) External time series. (c) Previous subsequence.

Fig. 10. Different references of normality.

Dissimilarity-based methods

- **Same time series**: considers the same time series as the reference.
  - Clustering techniques commonly use this to create clusters.
- **External time series**: rely on an external time series as a reference, assuming it has been generated by the same underlying process.
  - These series can be non-outliers subsequences of the same time series.
- **Previous subsequence**: only uses the previous adjacent non-overlapping windows as the reference.
  - A prediction model-based method.

Prediction Model-based methods

- Assumes that **normality** is reflected by a **time series composed of past subsequences**.
  - The intention is to **build a model** that captures the dynamics of the series and thus make future predictions. Subsequences far from those predictions are flagged as outliers.
- Any method for point-wise prediction model-based method can be used if adapted to work with subsequences.

Frequency-based methods

- Using also one of the references of normality mentioned in Fig. 10, a subsequence is an **outlier if it does not appear as frequently as expected**.
- The methods only work with a discretized time series.
  - Due to the difficulty of finding two exact real-valued subsequences in a time series when counting the frequencies.

Information Theory methods

- They assume that a subsequence that occurs frequently is less surprising and thus carries less information than a **rare subsequence**.
    - Their aim is to **find infrequent but still repetitive subsequences**, using the same time series as the reference of normality.
    - Very related to the frequency-based methods.

References

- Blázquez-García, Ane & Conde, Angel & Mori, Usue & Lozano, Jose. (2020). A review on outlier/anomaly detection in time series data.

Point Novelty and Outlier Detection with scikit-learn

Novelty vs Outlier

- **Outlier detection**: The training data contains outliers which are defined as observations that are far from the others.
    - Outlier detection estimators thus try to fit the regions where the training data is the most concentrated, ignoring the deviant observations.
- **Novelty detection**: The training data is not polluted by outliers and we are interested in detecting whether a new observation is an outlier.
    - In this context an outlier is also called a novelty.

- **Both** are used for anomaly detection.

Novelty Detection with One Class SVM

- *Is the new observation so different from the others that we doubt it's regular?*
  - This is the question addressed by the novelty detection tools and methods.
- In general, it is about to **learn a rough, close frontier** delimiting the contour of the initial observations distribution.
  - If new observations lay outside the frontier, we can say that they are abnormal.

Novelty Detection with One Class SVM

- The **One-Class SVM** has been introduced for that purpose.
  - It requires the choice of a kernel (usually RBF) and a scalar parameter to define a frontier.
  - The *nu* parameter, also known as the margin of the One-Class SVM, corresponds to the probability of finding a new, but regular, observation outside the frontier.

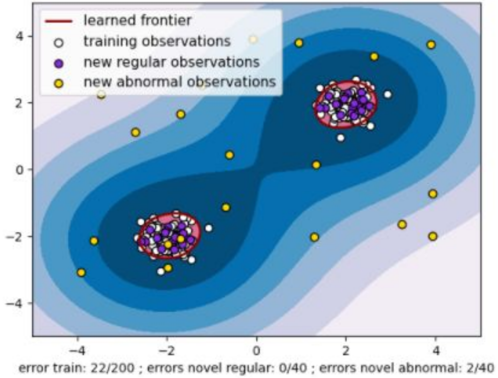


Fig. 11. Novelty detection.

Outlier Detection by fitting an Elliptic Envelope

- It assumes that data comes from a **known distribution** (e.g. Gaussian).
- Then fits a robust covariance estimate to the data, and thus **fits an ellipse to the central data points**, ignoring points outside the central mode.

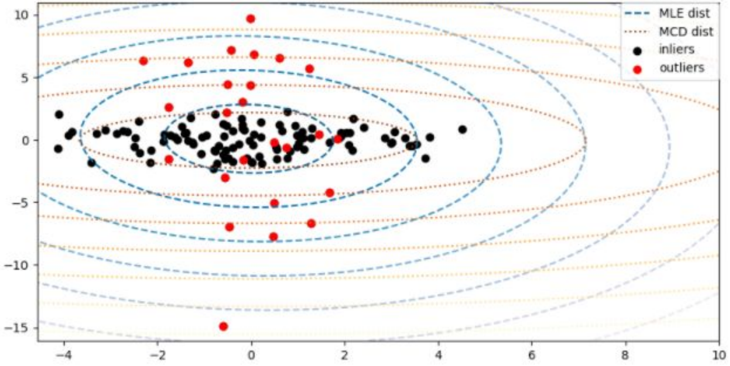


Fig. 12. Mahalanobis distances of a contaminated dataset.

Outlier Detection with Isolation Forest

- The Isolation Forest **‘isolates’ observations** by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature.
- The **path length** from the root node to the terminating node is a measure of normality and our decision function.
  - When a forest of random trees collectively produce <u>shorter path lengths</u> for particular samples, they <u>are highly likely to be anomalies</u>.

Outlier Detection with Isolation Forest

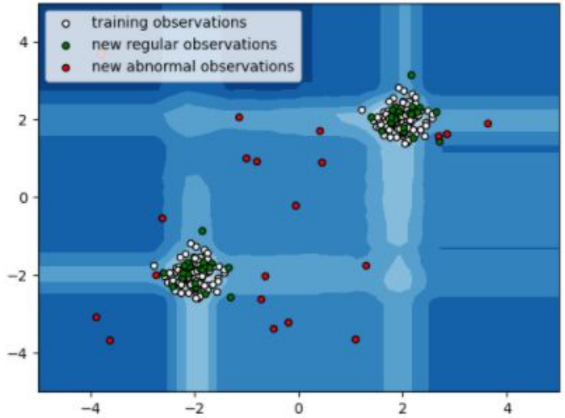


Fig. 13. Isolation Forest.

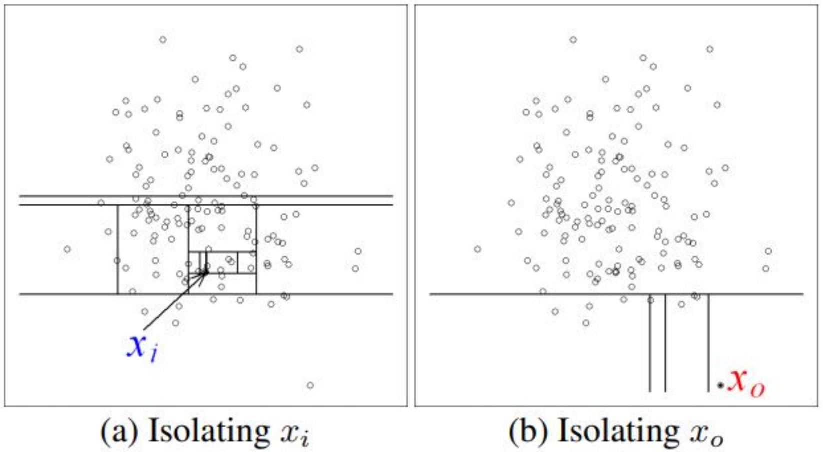


Fig. 14. Steps for isolating an inlier (a) and an outlier (b)

Outlier Detection with Local Outlier Factor (LOF)

- It measures the **local density deviation** of a given data point with respect to its neighbors.
    - The idea is to detect the samples that have a substantially lower density than their neighbors.
- In practice the local density is obtained from the **k-nearest neighbors**.
    - The LOF score of an observation is equal to the ratio of the average local density of his k-nearest neighbors, and its own local density.

Outlier Detection with Local Outlier Factor (LOF)

- A normal instance is expected to have a local density similar to that of its neighbors.
- **Abnormal data** are expected to have **much smaller local density**.

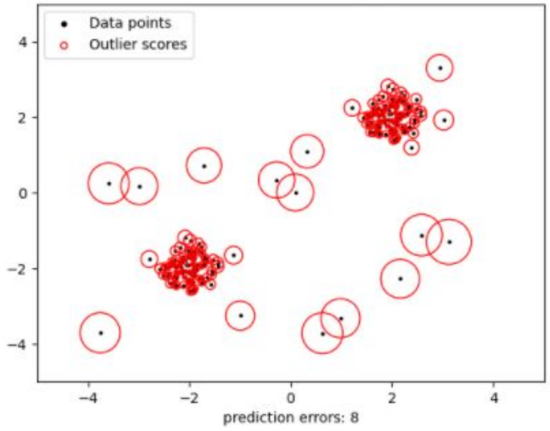


Fig. 15. Local Outlier Factor.

References

- Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
- F. T. Liu, K. M. Ting and Z. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422, doi: 10.1109/ICDM.2008.17.