

Weekly summary

Monday, August 12

Initiating

Tuesday, August 13

Independent component analysis (ICA) - a computational method for separating a multivariate signal into additive subcomponents. This is done by assuming that the subcomponents are non-Gaussian signals and that they are statistically independent from each other. FastICA algorithm.

Kernel functions

Nearest neighbour based methods - Expensive for multivariate data

Wednesday, August 14

Evaluation of techniques Use ROC-curve to represent trade-off. Measure area under the curve. (False positives/Detection rate) Src: Outlier Detection Techniques for Wireless Sensor Networks: A Survey

Challenges "The detection of anomalies in multivariate time series is more challenging for several reasons. First, it is difficult to establish a concise definition of an anomaly. Analogous to univariate time series, some anomalies may correspond to abnormally high (or low) values or unusual subsequences in one or more time series. In addition, the multivariate anomalies may correspond to unexpected changes in the relationships among a set of variables. For example, the time series for vegetation cover at mid- latitude locations in the United States typically varies in a 12-month cycle, peaking during the warm summer months and dropping to its minimum during the cold winter. Ecosystem disturbances such as wildfire and drought can be potentially detected based on the unusually low values of vegetation cover observed in the summer. Such anomalies are considered to be *local* (as opposed to global anomalies) since their values are abnormally low when compared to the average values observed during the warm summer months. Second, the performance of a multivariate anomaly detection algorithm is highly susceptible to the presence of noise in one or more time series. Therefore, a multivariate anomaly detection algorithm must be robust to noisy measurements in the time series data in order to increase its detection rate and to reduce its false alarm rate." Src: Detection and Characterization of Anomalies in Multivariate Time Series

Classification of detection techniques

1. Statistical

- Parametric

- Non-parametric
- 2. Nearest neighbour-based
- 3. Clustering
 - k-means
- 4. Classification
- 5. Spectral decomposition
-
- 6. Local outlier factor (LOF)?

Src: Outlier Detection Techniques for Wireless Sensor Networks: A Survey

In layman terms PCA helps to compress data and ICA helps to separate data.

”Change can occur slowly over time or abruptly, Outliers in a time series are either contextual (9.2) or collective (9.3) anomalies. Outliers are contextual when the values at specific time stamps suddenly change with respect to their temporally adjacent values, whereas outlier are collective when entire time series or large subsequences within a time series have unusual shapes.” Outlier analysis

FrÅ¥gor: Hur kan metoderna generaliseras? Unsupervised/supervised? Typ av tidsserie spelar stor roll fÅ¶r metodval. Online vs offline?

Thursday, August 15

Åverblick av metoder fÅ¶r anomalidetektion <https://blog.statsbot.co/time-series-anomaly-detection-algorithms-1cef5519aef2>

Sammanfattning av denna:

Types of anomalies

*Imagine you track users at your website and see an unexpected growth of users in a short period of time that looks like a spike. These types of anomalies are usually called **additive outliers**. Another example with the website is when your server goes down and you see zero or a really low number of users for some short period of time. These types of anomalies are usually classified as **temporal changes**. In the case that you deal with some conversion funnel, there could be a drop in a conversion rate. If this happens, the target metric usually doesnât change the shape of a signal, but rather its total value for a period. These types of changes are usually called **level shifts** or **seasonal level shifts** depending on the character of the change.*

Basically, an anomaly detection algorithm should either label each time point with anomaly/not anomaly, or forecast a signal for some point and test if this point value varies from the forecasted enough to deem it as an anomaly.

Methods

STL decomposition by Loess *This technique gives you an ability to split your time series signal into three parts: seasonal, trend and residue. If you analyze deviation of residue and introduce some threshold for it, you'll get an anomaly detection algorithm.*

Classification and regression trees (CART) *First, you can use supervised learning to teach trees to classify anomaly and non-anomaly data points. In order to do that you'd need to have labeled anomaly data points. The second approach is to use unsupervised learning to teach CART to predict the next data point in your series and have some confidence interval or prediction error as in the case of the STL decomposition approach. You can check if your data point lies inside or outside the confidence interval using Generalized ESD test, or Grubbs's test.*

ARIMA

Exponential smoothing

Neural networks *As in the case of CART, you have two ways to apply neural networks: supervised and unsupervised learning. As we're working with time series, the most suitable type of neural network is LSTM. This type of Recurrent Neural Network, if properly built, will allow you to model the most sophisticated dependencies in your time series as well as advanced seasonality dependencies. This approach can also be very helpful if you have multiple time series coupled with each other.*

<https://www.sciencedirect.com/science/article/pii/S0925231217309864> <https://hackernoon.com/how-far-have-we-gotten-in-time-series-prediction-from-rnn-to-lstm-5c9db64eb805> <https://www.elen.ucl.ac.be/Proceedings/esas56.pdf> <https://arxiv.org/ftp/arxiv/papers/1812/1812.00890.pdf>

Friday, August 16