Departamento de
Ingeniería de la Información
y las Comunicaciones

UNIVERSIDAD DE
MURCIA

Computer Sciences Faculty, University of Murcia

A combination of multi and univariate anomaly detection in urban irrigation systems

EDiS' 2022 : 3rd international conference on Embedded & Distributed Systems

Aurora González-Vidal, Jesús Fernández-García, Antonio F. Skarmeta
aurora.gonzalez2@um.es

Content

Introduction and Motivation

Background and related work

Methodology and experiments

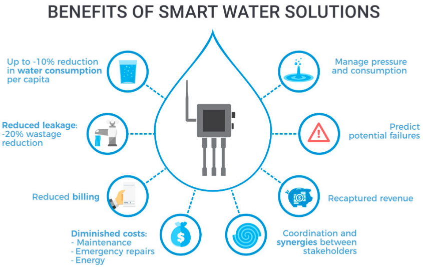
Conclusions

Introduction and Motivation

Introduction and Motivation

Water scarcity

Water scarcity is a global concern that requires a more intelligent way to manage the water network


BENEFITS OF SMART WATER SOLUTIONS
Up to -10% reduction in water consumption per capita
Reduced leakage: -20% wastage reduction
Reduced billing
Diminished costs: - Maintenance - Emergency repairs - Energy
Coordination and synergies between stakeholders
Recaptured revenue
Predict potential failures
Manage pressure and consumption

Automated and intelligent decision support systems that employ data analytic advances to analyze in real-time the anomalies and problems that appear in water usage are needed

Image:
`https://www.libelium.com/libeliumworld/quick-report-smart-water-iot-solutions-to-fight-against-climate-change-and-scarcity/`

Introduction and Motivation
Aim

Smart water meters periodically collect measurements of water consumption: multivariate time series

Anomaly detection purposes:

- Water quality
- Changes in consumption revealing leaks, device/meter failer, illegal water use, warning situations

Problem: anomalies are rarely anotated → A methodology based on unsupervised methods is needed

Background and related work

Background and related work

Types of oultiers

- Kind of input data
  - A univariate time serie $X = \{x_t\}_{t \in T}$ isan ordered set of real valued observations, where each observation is recorded at a specific time t
  - A multivariate time series $\mathbf{X} = \{\mathbf{x}_t\}_{t \in T}$ is an ordered set of k-dimensional vectors, each of which is recorded at a specific time $t$and consists of k real-valued observations, $x_t = (x_{1t}, ..., x_{kt})$.
- Kind of outlier
  - An point outlier x_t it behaves differently than what it was expectedas follows: $|x_t - \hat{x}_t| > \delta$
  - Pattern-wise outliers are a set of consecutive points in time whose joint behaviour is unusual: basic shapelet, seasonality changes and trend alternations

Background and related work

Related work - anomaly detection in water related scenarios

- Water usage anomalies in households using DBSCAN and K-means
- Punctual anomalous water consumption using STL decomposition and Seasonal Hybrid ESD Test
- Manually spike, jump and drift outliers comparing multivariate and univariate supervised methods

Existent approaches normally chose a strategy that is either multivariate or univariate and therefore, they cannot take advantage of the complex and dynamic behaviour present in water systems

Methodology and experiments

Methodology and experiments

Dataset clearning

The water consumption of parks in the city of Murcia, Spain is provided by EMUASA[1], the Municipal Water and Sanitation Company.

The water meter measurements are present in different granularities (minutely, hourly, daily, and even weekly sometimes). Most water meters have data between the 29th of September, 2017, and 25 of July, 2021.

Challenges:

- Negative water consumption
- Change of water meter
- Different starting and ending days

After cleaning, data was homogenously aggregated every 15 minutes and we count on 283 time series.

Methodology and experiments

Methodology

The data is huge in volume and dimension → Univariate methods are not feasible → We used a multivariate technique that will tell us for a group of time series, where there has been an anomaly
Methodology:

- Clustering the parks according to their water consumption,
- Applying multivariate anomaly detection in each cluster for identifying generally anomalous dates, and
- Applying univariate anomaly detection for each series in a range close to the previously-detected anomalous dates.

Methodology and experiments

Clustering

Hierarchical clustering was chosen. Distances - Let X_n and Y_n be two time series:

- Euclidean:

$$d_{L_2}\left(\boldsymbol{X}_n, \boldsymbol{Y}_n\right) = \left(\sum_{t=1}^{n} \left(X_i - Y_i\right)^2\right)^{1/2}$$

- Fréchet:

$$d_F\left(\boldsymbol{X}_n, \boldsymbol{Y}_n\right) = \min_{r \in M} \left(\max_{i=1,\dots,m} \left|X_{a_i} - Y_{b_i}\right|\right)$$

- DTWARP:

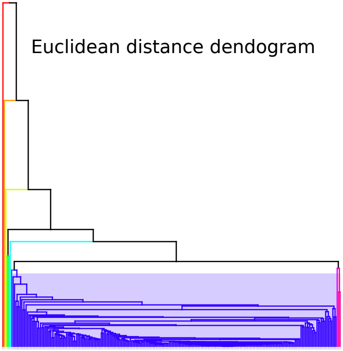
$$d_{DTWARP}\left(\boldsymbol{X}_n, \boldsymbol{Y}_n\right) = \min_{r \in M} \left(\sum_{i=1,\dots,m} \left|X_{a_i} - Y_{b_i}\right|\right)$$

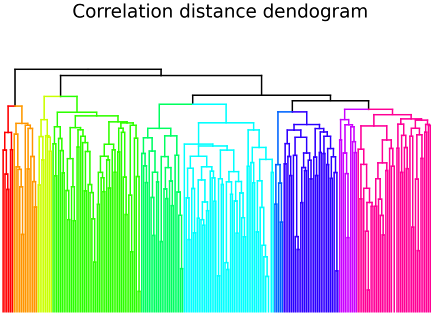
- Correlation-based:

$$\mathrm{COR}\left(\boldsymbol{X}_n, \boldsymbol{Y}_n\right) = \frac{\sum_{t=1}^{n} \left(X_t - \overline{\boldsymbol{X}}_n\right)\left(Y_t - \overline{\boldsymbol{Y}}_n\right)}{\sqrt{\sum_{t=1}^{n} \left(X_n - \overline{\boldsymbol{X}}_n\right)^2} \sqrt{\sum_{t=1}^{n} \left(Y_t - \overline{\boldsymbol{Y}}_{Tn}\right)^2}},$$

where M is the set of all possible sequences of m pairs of order-preserving observations and $\overline{\boldsymbol{Y}}_n$ and $\overline{\boldsymbol{X}}_n$ are the average of their corresponding series.

Manual selection of 10 groups. Euclidean $\sim$ Fréchet and DTW $\sim$ correlation-based distances





Methodology and experiments

Mutlivariate Anomaly Detection

For the multivariate anomaly detection we have used the Vector autoRegression (VAR) model. The formula of the model is as follows:

$$\mathbf{y}_t = \alpha_t + \Phi_1 \mathbf{y}_{t-1} + ... + \Phi \mathbf{y}_{t-p} + \epsilon_t$$

y_{t-i} is the “ith lag" of y_t. α is a k-vector of constants serving as the intercept of the model. Φ_i is a time-invariant (k × k)-matrix and ϵ_t is a k-vector of error terms.

As an example the VAR(1) of 3 time series denoted by $y_{t,1}$,$y_{t,2}$, and $y_{t,3}$ is as follows:

$$\begin{aligned}
y_{t,1} &= \alpha_1 + \phi_{11} y_{t-1,1} + \phi_{12} y_{t-1,2} + \phi_{13} y_{t-1,3} + \epsilon_{t,1} \\
y_{t,2} &= \alpha_2 + \phi_{21} y_{t-1,1} + \phi_{22} y_{t-1,2} + \phi_{23} y_{t-1,3} + \epsilon_{t,2} \\
y_{t,3} &= \alpha_3 + \phi_{31} y_{t-1,1} + \phi_{32} y_{t-1,2} + \phi_{33} y_{t-1,3} + \epsilon_{t,3}
\end{aligned}$$

Methodology and experiments
VAR

VAR for anomaly detection:

$$SE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2,$$

where y_i are the real values and $\hat{y}_i$ are the VAR predictions for each observation i. Then we define a threshold that is:

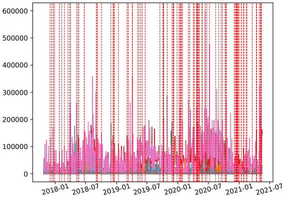
$$threshold = \frac{1}{n} SE + sd(SE),$$

where sd is the standard deviation. If $SE \geq$ threshold $\rightarrow$ ANOMALY

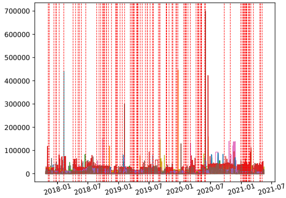
Methodology and experiments

VAR

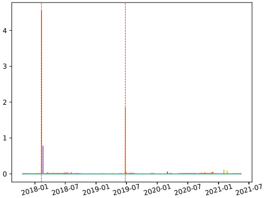
Cluster 3



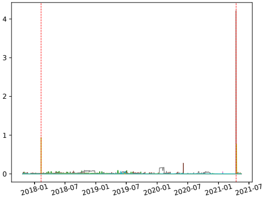
Cluster 4



Cluster 5



Cluster 9



Methodology and experiments

Univariate anomaly detection

Let z_t be an ARIMA model, then it is assumed that our series y_t, which contains m outliers can be described as:

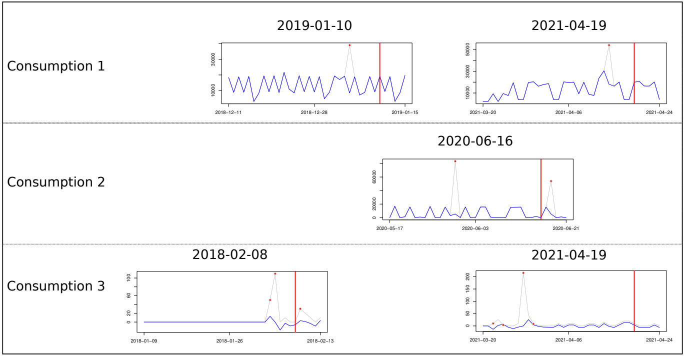
$$y_t = z_t + m \sum_{j=1}^{m} w_j L_j(B) I_{t_j},$$

- aditive outliers
- level shifts
- temporary changes
- innovative outliers
- seasonal level shift

For further information about how each of the outliers is defined, please check the implementation manual of the R package *tsoutliers* or the original paper "Chen, C., & Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association, 88(421), 284-297"

Methodology and experiments

Univariate anomaly detection



Conclusions

Conclusions / future work

- We have developed a methodology for unsupervised anomaly detection in IoT deployments
- We captured general knowledge present in several series through mutlviariate analysis and then searched for the specific anomalies using univariate analysis
- Our method is applied for searching anomalies in the irrigation of urban parks, but it is general enough for its use in greater agricultural set-ups

Departamento de
Ingeniería de la Información
y las Comunicaciones

UNIVERSIDAD DE
MURCIA

Questions?

Find this presentation at: `github.com/auroragonzalez/presentations`

Computer Sciences Faculty, University of Murcia

A combination of multi and univariate anomaly detection in urban irrigation systems

EDiS' 2022 : 3rd international conference on Embedded & Distributed Systems

Aurora González-Vidal, Jesús Fernández-García, Antonio F. Skarmeta
aurora.gonzalez2@um.es