Computer Sciences Faculty
University of Murcia

IoT for Water Management: Towards Intelligent Anomaly Detection

2019 IEEE WF-IoT. TP1-8: IoT Services and Applications for Verticals

Aurora González-Vidal, Jesús Cuenca-Jara, Antonio F. Skarmeta
{aurora.gonzalez2, jesus.cuenca1, skarmeta}@um.es

April 18th, 2019

Content

Introduction

Context

- Yearly >32 billion m^3 of treated water are lost
- Water demand dynamics make it difficult to manage water supply
- Smarter ways: IoT + data analytics
- Smart water meters collect measurements of water consumption, opening up a wealth of opportunities for water providers when empowered by data mining techniques.

Anomalies in water consumption

Anomalies in water consumption time series lead to

- Reveal leaks
- Reveal device/meter failures
- Detect illegal water use
- Detect warning situations
- Allow the provision of personalized feedback to customer

Objective

- To develop a methodology in order to detect **anomaly consumption in water systems**
- Application: evaluation (testing the effectiveness) in a real water system
- Comparison of two approaches: ARIMA-based framework and HOT-SAX-based framework

IoTCrawler

Related work

Related work

Several techniques have already been used for anomaly detection in the urban water process:

- **Change-point** methods
- Probabilistic **outlier detection** has been explored in water management scenarios.
- Heatwave events in temperature time series can be considered anomalies. In the litterature, a multiresolution quantile approach that extends a variation of the Symbolic Aggregate Approximation (SAX) that we are also going to use is proposed to find heatwave events.

Methodology

Scheme

- **Objective:** discover the weak and strong points of methods from different nature to find anomalies in water consumption time series
- **Two-steps scheme**
  1. Extract outliers and abnormal patterns using the individual time series properties of the data
  2. Try to classify them thanks to the annotated series.

Step 1 - ARIMA

ARIMA based algorithm

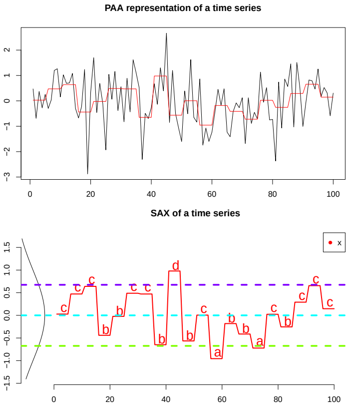
$$y_t = z_t + \sum_{j=1}^{m} w_j L_j(B) I_{tj}$$

z_t is an ARIMA process

- Five types of **outliers** are considered:
    1. “AO” additive outliers
    2. “LS” level shifts
    3. “TC” temporary changes
    4. “IO” innovative outliers
    5. “SLS” seasonal level shifts

Step 1 - HOT-SAX

- Symbolic Aggregate Approximation (SAX)
  1. Normalization
  2. Discretization: Piecewise Aggregate Approximation (PAA)
  3. Symbolization (alphabet)
- Heuristically Order Time series (HOT): find the less frequent words in the time series –> anomalies



Step 2 - ARIMA

Association Rule Learning for ARIMA framework

- **Objective:** differentiate which types of outliers correspond to anomalous behaviors and which correspond to normal fluctuations in consumption
- An **Association Rule Learning Algorithm** was applied to these sets of outliers with the aim of studying if the type of outliers detected by the algorithm is representative for each class.

Step 2 - ARIMA

- In order to carry out a more exhaustive study, the Association Rule Learning Algorithm was applied on **different size sets**:
  1. Taking only the outlier closest to the breakdown
  2. Taking the two closest
  3. Taking the three closest
  4. etc.

Step 2 - HOT-SAX

Subseries classification using Random Forest for HOT- SAX discordances

- **Objective:** discriminate between subseries that indicate a future anomaly and subseries that do not after applying HOT-SAX algorithm
  1. **Separate series** that lead to anomalies and series which does not lead to anomalies and apply the algorithm
  2. After that, we apply a **random forest** classifier in order to fit a model that discriminates between subseries

Metrics

- Contingency Table

	Anomaly	Not Anomaly
Anomaly predicted	True positive (TP)	False positive (FP)
Not Anomaly predicted	False negative (FN)	True negative (TN)

- Sensitivity and Specificity

$$Sensitivity = \frac{TP}{TP + FN}$$

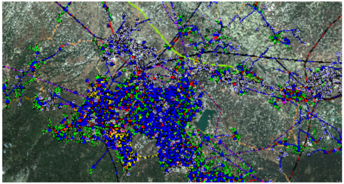
$$Specificity = \frac{TN}{TN + FP}$$

Use case and experiment results

Data

- Data from **40000 smart meters** that mainly belong to factories and buildings
- Every data point is gathered with a timestamp, that is, it has the form of a time series.
- **30 time series** anotated

SCADA real time pipe information



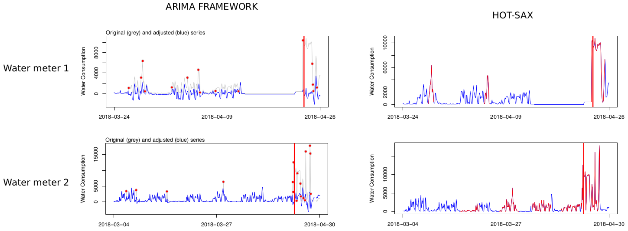
Results - Detection of anomalies (Step 1)

Accuracy detecting anomalies:

- ARIMA-framework: **90 %**
- HOT-SAX: **80 %**

However, both approaches present a high rate of false positives

Results on 2 smart meters



Results Classification of anomalies (Step 2)

- ARIMA + Association Rules: Inconclusive
- HOT-SAX + Random Forest:

Table: Contingency Table Results

	Anomaly	Not Anomaly
Anomaly predicted	**12**	9
Not Anomaly predicted	2	**23**

$$Sensitivity = \mathbf{86\%}$$

$$Specificity = \mathbf{72\%}$$

Conclusions and future work

Conclusions

- This is the **first time** that a combination between univariate anomaly detection techniques and further analysis is used for discarding false anomalies in water consumption
- This is a **step forward** towards automatic water management in smart cities. In many utility companies, anomaly detection is either neglected or done by a technician who normally is unable to check all smart meters

Future Work

- Select other algorithms based on a machine learning approach
- Design an experiment which tests several anomaly detection models and finds the best one through expert validation.
- Elaborate a friendly graphic user interface
- Analyze and develop retrofit techniques
- Use Big Data frameworks in order to facilitate the analysis on real time.

Highlighted references

- C. Chen, L.-M. Liu, Joint estimation of model parameters and outlier effects in time series, Journal of the American Statistical Association 88 (421) (1993) 284–297.
- J. L. de Lacalle, tsoutliers: Detection of Outliers in Time Series, r package version 0.6-6 (2017). URL https://CRAN.R-project.org/package=tsoutliers
- Keogh, Eamonn, Jessica Lin, and Ada Fu. "Hot sax: Efficiently finding the most unusual time series subsequence." Fifth IEEE International Conference on Data Mining (ICDM'05). Ieee, 2005.

Questions?

Find this presentation at:

`github.com/auroragonzalez/presentations`

Computer Sciences Faculty
University of Murcia

IoT for Water Management: Towards Intelligent Anomaly Detection

2019 IEEE WF-IoT. TP1-8: IoT Services and Applications for Verticals

Aurora González-Vidal, Jesús Cuenca-Jara, Antonio F. Skarmeta
{aurora.gonzalez2, jesus.cuenca1, skarmeta}@um.es