MACHINE LEARNING METHODS FOR ANOMALY DETECTION:

Examples from Dam and Levee Condition Monitoring

Valeria Krzhizhanovskaya
University of Amsterdam, The Netherlands
ITMO University, Russia

V.Krzhizhanovskaya@uva.nl
OUTLINE

- IoT, Big Data, Data Science
- Data-driven modelling
- Machine learning
- Anomaly detection

Example:
   dam and levee condition monitoring
Big Nonsense: The end of scientific thinking is near...

Peter M.A. Sloot

The end of scientific thinking is near...

P.M.A. Sloot.

“The end of scientific thinking is near…”


World Scientific Publishing Co. Pte. Ltd.

Big Nonsense: the end of scientific thinking is near...

Peter M.A. Sloot

‘The real purpose of the scientific method is to discover that Nature hasn’t misled you into thinking you know something you don’t actually know.’

Robert M. Pirsig, 1974

Astronomers of the Maya civilisation and astronomers of the Babylonian civilisation were brilliant in predicting astronomical events. For instance, from meticulous observations of the Sun, Moon, Venus and Jupiter they were able to predict with astonishing accuracy the 584-day cycle of Venus or the details of the celestial track of Jupiter [1]. Yet they had no clue about our heliocentric solar system, they believed that the earth was flat and they were completely ignorant of the real movement of stars and planets while being convinced that the sky was supported by four jaguars, each holding up a corner of the sky. If we would be sent back in time and speak to them about the planets orbiting the sun they would laugh at us and challenge us to come with the accurate predictions that they are able to make. With all our knowledge, but without thousands of years of technological development, we would not be able to come close to any of their predictions. So being laughed at would be a small punishment, more likely we would be ritually slaughtered...

This is how I feel, time and again when one or another data fetishist tries to convince me of the importance of big data over proper scientific inference and methods.

We are on the edge of a fundamental clash of scientific methods. If we do not solve this soon we risk being thrown back into pre-medieval scientific methods. This warning should reverberate strongly for us complexity- and system scientists, as we are slowly being drawn into data obscurantism.
How big is the Big Data?
Big data science is dangerous but it is all around us!

- Data science ➔
  - new methods and tools
  - to derive mathematical models from real-life observations
  - to detect anomalies
  - to predict the future
DATA SCIENCE + COMPUTATIONAL SCIENCE
=
?
UrbanFlood.eu

Physics-based simulation
Gate vibrations
Analytical
Machine learning
**Flood Barrier Gate Vibrations**

**Simulation**

**Experiment**

- **Data Acquisition**: $h_1(t)$, $h_2(t)$, $x_i'(t)$, $a(t)$
- **Primary Data Analysis**
  - $f_{dom}$, $X_{dom,l}$, $\Delta h$, $\dot{a}$, $A_{corr}$
- **Machine Learning**
  - Prediction of hydraulic parameters
- **Physics Simulations**
  - Computational Fluid-Structure Interaction
C.D. Erdbrink, V.V. Krzhizhanovskaya, P.M.A. Sloot, "Reducing cross-flow vibrations of underflow gates: Experiments and numerical studies", Journal of Fluids and Structures, V. 50, 2014, pp. 25-48, doi: 10.1016/j.jfluidstructs.2014.06.010

C.D. Erdbrink, V.V. Krzhizhanovskaya, "Differential evolution for system identification of self-excited vibrations", Journal of Computational Science, V. 10, 2015, pp. 360-369. doi: 10.1016/j.jocs.2015.03.004
DATA-DRIVEN LEVEE MODELING AND ANOMALY DETECTION
2D full-scale simulation of porous flow & structural dynamics
**INPUT DATA**

- Cross section geometry
- Soil build-up
- Water level sensor signal
SIMULATION COMPONENTS (1)

- Porous flow:
  - Richards’ eq. for water flow in partially saturated soils
  - Van Genuchten model for water retention

\[ (C + \theta e S) \frac{\partial p}{\partial t} + \nabla \cdot [-K_s k_r \nabla (p + \rho g z)] = \frac{\partial \varepsilon}{\partial t} \]

FSI coupling terms
SIMULATION COMPONENTS (2)

- **Structural stability:**
  - Plane stress
  - Linear elastic soil
  - Mohr-Coulomb failure criterion

\[
\tau_f = c + \sigma_f \tan \phi
\]

FSI coupling terms

\[
\begin{align*}
\nabla \cdot \sigma &= -\nabla p + \rho_s g = 0 \\
\sigma &= \lambda \varepsilon E + 2\mu \varepsilon
\end{align*}
\]

unstable
VIRTUAL DIKE: BOSTON

Real layers

Initial settlement

Low tide

High tide
MODEL CALIBRATION BASED ON DATA

- Pressure head, mm
- 1E4 virtual
- E4 (K=1E-9m2)
- virtualE4 (K=1E-10m2)

- Pressure head, mm
- 1G2 virtual
- G2 (k=1E-9m2)
- virtualG2 (k=1E-10m2)
ANOMALY DETECTION APPROACH

Analytical redundancy

Physical redundancy (the same placement)

Physical redundancy (type of sensor)

Logical groups

Pre-processing
Feature Extraction

Cl_1

Cl_2

Cl_k

Decision support

Sensor measurements

Committee

Feature extraction

Confidence values

NORMAL BEHAVIOUR

ABNORMAL BEHAVIOUR
**PRE-PROCESSING**

The main idea: to apply methods, which have minimal number of adjustable parameters and require minimal information about signal

1) Wavelet denoising

![Wavelet Denoising Diagram](image)

\[ y \xrightarrow{\text{DWT}} \text{Coefficients} \xrightarrow{\text{Hard thresholding}} \text{iDWT} \xrightarrow{x} \]

2) Spectrum Singular Analysis (SSA)

\[ y \xrightarrow{\text{SSA}} y = \sum_{n=1}^{N} C_n \xrightarrow{\text{Take only } c \text{ which corresponds to maximal eigenvalues}} x \]

3) Hodrick-Prescott filter

\[ \min \sum_{i=1}^{T} (y_i - x_i)^2 + \lambda \sum_{t=2}^{T-1} [(x_{i+1} - x_i) - (x_i - x_{i-1})]^2, \lambda - \text{smoothing parameter} \]

4) L1 trend filtering

\[ \min \sum_{i=1}^{T} (y_i - x_i)^2 + \lambda \sum_{i=2}^{T} |x_i - x_{i-1}|, \lambda - \text{smoothing parameter} \]

5) Moving average

\[ x_i = \frac{1}{N} \sum_{l=0}^{N-1} y_{i-l} \]

\[ y – \text{measurements} \]

\[ x – \text{estimation of a signal} \]
**DATA PRE-PROCESSING**

**Synthesized data**

<table>
<thead>
<tr>
<th></th>
<th>Wavelet</th>
<th>l1</th>
<th>hp</th>
<th>SSA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.23</td>
<td>0.24</td>
<td>0.49</td>
<td>0.50</td>
<td>1.18</td>
</tr>
</tbody>
</table>

**Smoothing of ‘jumps’**

**Smoothing of outliers**
ONE-SIDE CLASSIFICATION

1) Hypercube

2) Parzen window

3) Neural Clouds

--- ABNORMAL BEHAVIOUR

-- NORMAL BEHAVIOUR
SIMULATED SENSOR DATA
ANOMALY DETECTION IN ARTIFICIAL DATA

First principal strain

X deformation

Confidence value of dike normal behaviour

Stability factor

Detection of anomaly

"Local failure"
COMBINING AI WITH VIRTUAL DIKE
AI COMPONENT AS PART OF THE URBANFLOOD EARLY WARNING SYSTEM

Pyayt et al. 2015 "Combining Data-Driven Methods with Finite Element Analysis for Flood Early Warning Systems", Procedia Computer Science, V. 51, 2015, pp. 2347-2356, doi: 10.1016/j.procs.2015.05.404


LATEST WORK ON MONITORING LEVEES

Photo from www.cnn.com, South Carolina flooding: Dams breached, more trouble ahead, Oct 7, 2015 Update.
Earth Dams and Levees

- Internal Erosion
- Worldwide Problem

Approach: Non-intrusive(!) monitoring

OUR APPROACH

Data Collection
- Experimental Levees
- Crack-Box Testbed
- Piping Experiment
- Passive Seismic Data

Preprocessing
- Wavelet Denoising
- Spectral Frames
- Feature Extraction
- Standardization
- Automatic Feature Selection

Machine Learning
- Classification
- Anomaly Detection

Results
- Visualization
- Validation
PASSIVE SEISMIC DATA

- Crack-Box & Piping Experiments
  - Internal Erosion and Cracking
  - Geophones
  - 4,140 Seconds of Data
DATA PREPROCESSING

1. Wavelet de-noising: MODWT, Haar level 3

2. Spectral frames (1, 2, 3, 5, and 10-second frames)

3. Feature extraction
3. Feature extraction
4. Standardization (zero mean, unit STD)

+ Flatness, Kurtosis, Irregularity, Skewness
5. AUTOMATIC FEATURE SELECTION

- Pre-processed Input Data X
  - iFeat <= numFeat
    - combos = nChoosek (n = numFeat, k = iFeat)
      - iCombos <= numCombos
        - Set Features
          - Train OCSVM & Perform 10-fold Cross Validation
            - Calculate Results Metrics
              - CurrentFscore > MaxFscore
                - Store MaxFscore
                  - Visualization
## Selected Features

### Crack

<table>
<thead>
<tr>
<th>Feature</th>
<th>1 Second</th>
<th>2 Second</th>
<th>3 Second</th>
<th>5 Second</th>
<th>10 Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.957</td>
<td>0.961</td>
<td>0.965</td>
<td>0.949</td>
<td>0.986</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.973</td>
<td>0.976</td>
<td>0.978</td>
<td>0.968</td>
<td>0.991</td>
</tr>
<tr>
<td>Best # of features</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Selected features</td>
<td>RMS, SP</td>
<td>RMS, KU</td>
<td>RMS, KU</td>
<td>RMS, RG</td>
<td>RMS, FL</td>
</tr>
</tbody>
</table>

### Piping

<table>
<thead>
<tr>
<th>Feature</th>
<th>1 Second</th>
<th>2 Second</th>
<th>3 Second</th>
<th>5 Second</th>
<th>10 Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.940</td>
<td>0.950</td>
<td>0.935</td>
<td>0.919</td>
<td>0.951</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.963</td>
<td>0.969</td>
<td>0.960</td>
<td>0.950</td>
<td>0.970</td>
</tr>
<tr>
<td>Best # of features</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Selected features</td>
<td>ZC, RMS</td>
<td>ZC, RMS</td>
<td>ZC, RMS</td>
<td>ZC, RMS, RG</td>
<td>ZC, RMS, RG</td>
</tr>
</tbody>
</table>

**Relieff filter method:**
- k-nearest neighbors per class
- weighted feature ranking

**Graph:**
- **Y-axis:** Weight [-1, 1]
- **X-axis:** Zerocross, Centroid, Spread, RMS, Rolloff, Flatness, Kurtosis, Irregularity, Skewness
- **Legend:**
  - 1 Sec
  - 2 Sec
  - 3 Sec
  - 5 Sec
  - 10 Sec
Clustering

Anomaly Detection
STAGES OF CLUSTERING

- Pattern Representation
- Distance Measure
- Grouping Membership
- Cluster Representation
- Validation Techniques
CLUSTERING ALGORITHMS

- K-Means Clustering
- Partitioning Around Medoids
- Fuzzy C-Means Clustering
- Gaussian Mixture Model
- Agglomerative Hierarchical Clustering
Cluster Representation

- Principal Component Analysis
  - K-means
  - K=3
  - 10-Second Frames
CLUSTER REPRESENTATION

- 3D Plot of Selected Features
  - K-means
  - K=3
  - 10-Second Frames
CLUSTER REPRESENTATION

- Time Series Cluster Color Overlay
  - K-means
  - K=3
  - 10-Second Frames
INTERNAL VALIDATION

- Silhouette Width
  - Range: [-1, 1]
  - Averaged over 100 Runs
  - 95% CI
  - 10-Second Frames
EXTERNAL VALIDATION

- Purity
  - Range: [0, 1]
  - Averaged over 100 Runs
  - Comparing all 5 Algorithms

<table>
<thead>
<tr>
<th>Frame Size</th>
<th>Obs.</th>
<th>KM</th>
<th>HC</th>
<th>GMM</th>
<th>PAM</th>
<th>FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Seconds</td>
<td>1380</td>
<td>0.830</td>
<td>0.827</td>
<td>0.715</td>
<td>0.825</td>
<td>0.826</td>
</tr>
<tr>
<td>5 Seconds</td>
<td>828</td>
<td>0.825</td>
<td>0.829</td>
<td>0.728</td>
<td>0.823</td>
<td>0.826</td>
</tr>
<tr>
<td>10 Seconds</td>
<td>414</td>
<td>0.831</td>
<td>0.838</td>
<td>0.765</td>
<td>0.829</td>
<td>0.831</td>
</tr>
</tbody>
</table>

4 out of 5 Algorithms Performed well (> 82%)

GMM is less accurate
TWO-CLASS SUPPORT VECTOR MACHINE

- Nine Features
- RBF Kernel
- Cross Validation
- 97%+ Accuracy
**Prediction Result: 97%+ Accuracy**

Confusion Matrix Sample:
- **TP**: True Positive
  - Hit
  - Predicted normal on normal data
- **FP**: False Positive
  - False alarm
  - Predicted anomaly on normal data
- **FN**: False Negative
  - Miss
  - Predicted normal on anomalous data
- **TN**: True Negative
  - Correct reject
  - Predicted anomaly on anomalous data

Confusion Matrix, Frame Size=3:
- 365
  - 79.3%
- 3
  - 0.7%
- 2
  - 0.4%
- 90
  - 19.6%
ONE-CLASS SUPPORT VECTOR MACHINE

- Two Features
- RBF Kernel
- Cross Validation
- 83%+ Accuracy


**PREDICTION RESULT: 83%+ ACCURACY**

![Confusion Matrix Sample](image)

**TP**
- True Positive
  - Hit
  - Predicted normal on normal data

**FP**
- False Positive
  - False alarm
  - Predicted anomaly on normal data

**FN**
- False Negative
  - Miss
  - Predicted normal on anomalous data

**TN**
- True Negative
  - Correct reject
  - Predicted anomaly on anomalous data

![Confusion Matrix, Frame Size=2](image)

- **TP (True Positive)**: 493
  - 71.4%
- **FP (False Positive)**: 56
  - 8.1%
- **FN (False Negative)**: 57
  - 8.3%
- **TN (True Negative)**: 84
  - 12.2%
AUTOMATIC FEATURE SELECTION

- Improved accuracy from 83% to over 91%
W. Belcher, T. Camp, V. V. Krzhizhanovskaya, "Detecting Erosion Events in Earth Dam and Levee Passive Seismic Data with Clustering", 2015, doi: 10.1109/ICMLA.2015.9

DATA-DRIVEN TRAFFIC SIMULATION
**DATA**

- NDW = National Data Warehouse for traffic information
- 40,000 sensors in the Netherlands

<table>
<thead>
<tr>
<th>Vehicle length categories, meters</th>
<th>Data types</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 1</strong></td>
<td><strong>Set 2</strong></td>
</tr>
<tr>
<td>1,85 – 2,4</td>
<td>... – 5,6</td>
</tr>
<tr>
<td>2,4 – 5,6</td>
<td></td>
</tr>
<tr>
<td>5,6 – 11,5</td>
<td>5,6 – 12,2</td>
</tr>
<tr>
<td>11,5 – 12,2</td>
<td></td>
</tr>
<tr>
<td>12,2 – ...</td>
<td>12,2 – ...</td>
</tr>
</tbody>
</table>

**Data types**

<table>
<thead>
<tr>
<th>Traffic data</th>
<th>Status data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic flow (intensity)</td>
<td>Road works</td>
</tr>
<tr>
<td>Speed</td>
<td>Incidents</td>
</tr>
<tr>
<td>Travel time</td>
<td>Bridge/dynamic lanes status</td>
</tr>
</tbody>
</table>
OTHER SOURCES OF DATA
BLACKOUT FRIDAY TRAFFIC

1. Airport Schiphol
2. Amsterdam Sloterwaard
3. Amsterdam Sloterdijk
4. Amsterdam Tuindorp Oostzaan
5. Amsterdam Noord
6. Amsterdam Watergraafsmeer
7. Amsterdam Rivierenbuurt

Graphs showing speed and flow with marked Blackout Friday and normal times.
FIRE (BANGALORE)
TRAFFIC MODEL DEVELOPMENT BASED ON PARTIAL DATA (St. Petersburg)
Valentin Melnikov

V.R. Melnikov, V.V. Krzhizhanovskaya, A.V. Boukhanovsky, P.M.A. Sloot, "Data-driven Modeling of Transportation Systems and Traffic Data Analysis During a Major Power Outage in the Netherlands", Procedia Computer Science, V. 66, 2015, pp. 336-345. doi: 10.1016/j.procs.2015.11.039

V.R. Melnikov, V.V. Krzhizhanovskaya, M.H. Lees, A.V. Boukhanovsky, "Data-driven Travel Demand Modelling and Agent-based Traffic Simulation in Amsterdam Urban Area", Procedia Computer Science, V. 80, 2016, pp. 2030-2041. doi: 10.1016/j.procs.2016.05.523
CONCLUSIONS

- Data Science
  + Computational Science
  + Computer Science

- Machine learning is great for
  - Model calibration
  - Anomaly detection

- Even 1-class SVM can give 91% accuracy

- More challenges and exciting time ahead!
Brace yourselves for an interesting ride!
Many thanks for your attention!

Valeria Krzhizhanovskaya

University of Amsterdam, The Netherlands
ITMO University, Russia

V.Krzhizhanovskaya@uva.nl