rapidminer

Anomaly Detection and Structural Analysis in Industrial Production Environments

David Arnu[1], Martin Atzmüller[2], Andreas Schmidt[3]

[1] RapidMiner GmbH, [2] Tillburg University, [3] University of Kassel

iDSC 2017 Salzburg, Austria

Heterogeneous Mass data

Big Data Analysis

Support System

Operator

Research Project FEE –

„Frühzeitige Erkennung und Entscheidungsunterstützung für kritische Situationen im Produktionsumfeld“

Objective: Operator Support functions

- Early Warnings
- Ad-hoc Analysis
- Decision Support

Approach: Integrated Analysis of all plant data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

FEE – Data and System Landscape

Engineering Data

Laboratory Data

Alarms & Events

Process Measurements

Asset Data

Operation Manuals

Digital Shift Book

Big Data Platform

Assistence Systems
- Early Warnings
- Anomaly Detection & Scenario Search
- Ad-hoc-Analysis & Prognosis

Early Warning

Interactive Usage

Learning from History

Production Plant

Operators

Anomaly Detection in Process Industry

- Very few critical events
- Drift of concepts (process parameters, system changes)
- Two data sources
  - sensor readings
  - alarm messages
- Two stakeholders
  - *Operator*: continuously monitoring the plant; has to react to sudden changes
  - *Process engineer*: monitors overall trends; long term observations

Industrial Plant - Alarm Patterns

- Set of assets (part of a plant)
- Each asset contains a set of measurements
- Value range is monitored to trigger alarms

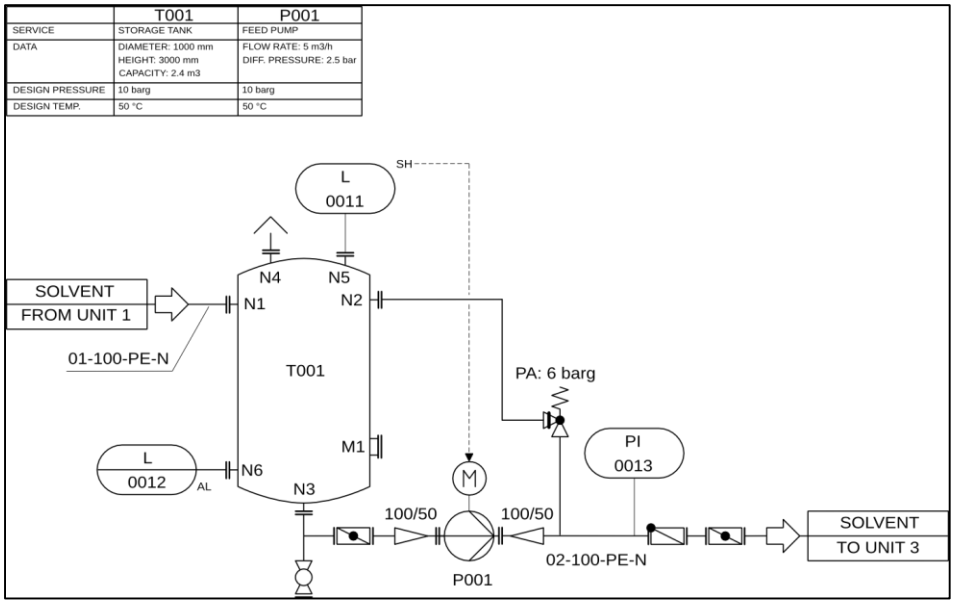
- Sequence of alarms grouped by asset
  - Snapshot of an abstract state of the plant
  - Model and compare these states

Industrial Plant - Alarm Patterns

- Set of assets (part of a plant)
- Each asset contains a set of measurements
- Value range is monitored to trigger alarms

- Sequence of alarms grouped by asset
  - Snapshot of an abstract state of the plant
  - Model and compare these states

Piping and Instrumentation Diagram (P&ID)



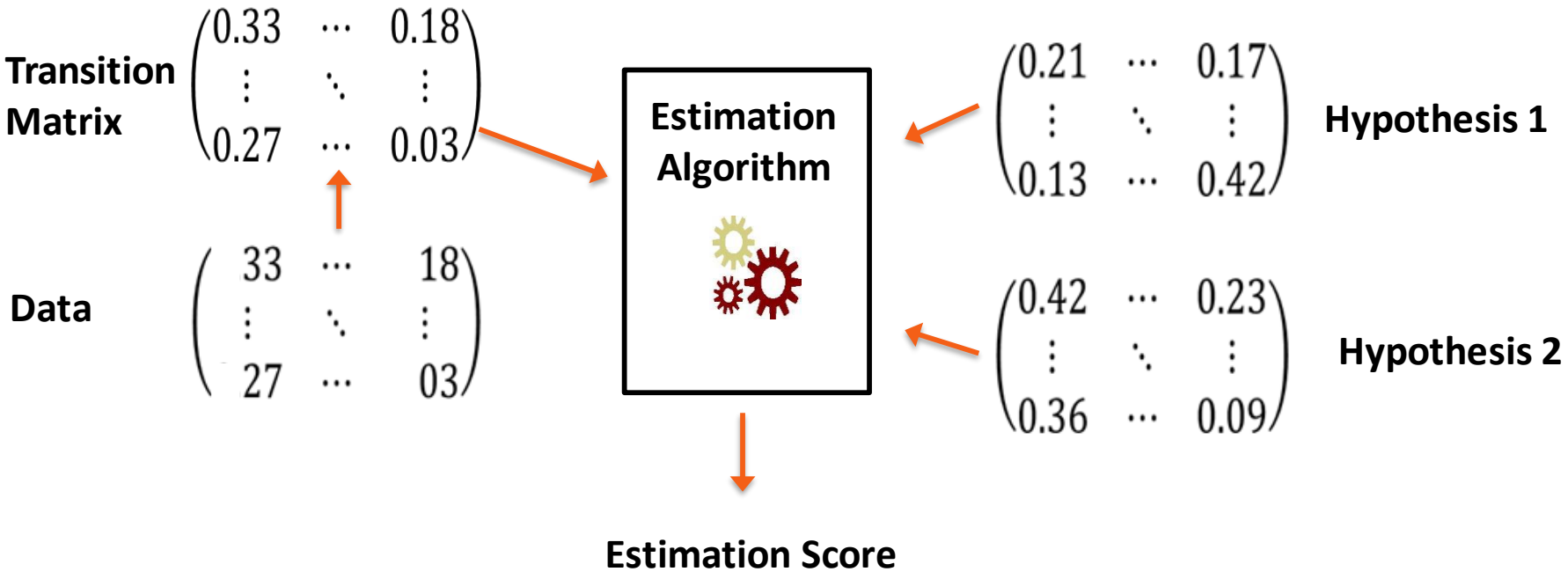
By Con-struct - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=18473227

HypGraphs - Graph-Based and Sequential Hypotheses

- First-Order Markov Chain modeling:
  Model transitions between different states

- Given a probability distribution on certain events
  (e.g., on alarms on different sub-parts of a plant)
  - Determine transition model
  - Collect transition matrix
  - Compare with hypotheses

HypGraphs (2)

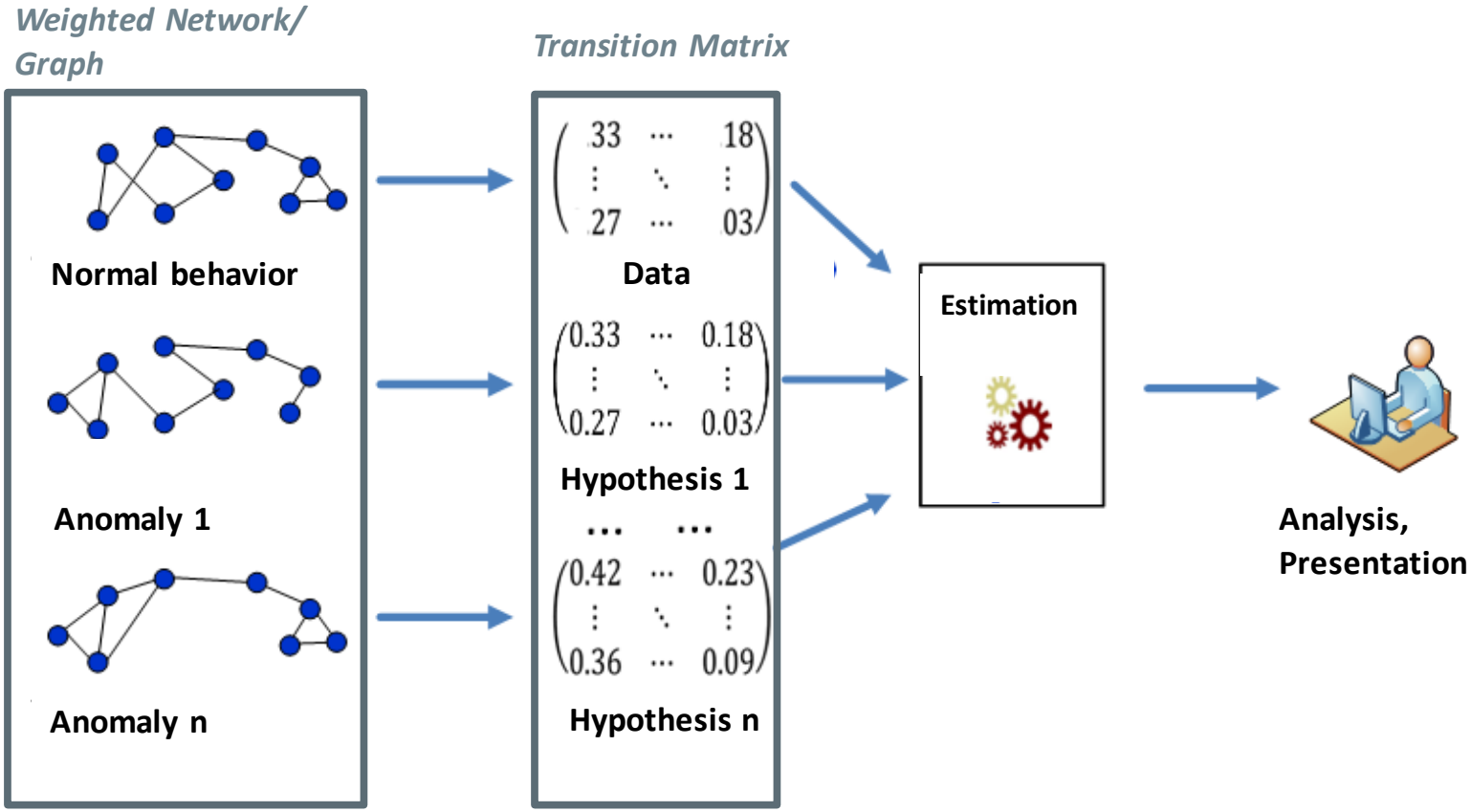

Transition Matrix
$\begin{pmatrix} 0.33 & \cdots & 0.18 \\ \vdots & \ddots & \vdots \\ 0.27 & \cdots & 0.03 \end{pmatrix}$
Data
$\begin{pmatrix} 33 & \cdots & 18 \\ \vdots & \ddots & \vdots \\ 27 & \cdots & 03 \end{pmatrix}$
Estimation Algorithm
$\begin{pmatrix} 0.21 & \cdots & 0.17 \\ \vdots & \ddots & \vdots \\ 0.13 & \cdots & 0.42 \end{pmatrix}$
Hypothesis 1
$\begin{pmatrix} 0.42 & \cdots & 0.23 \\ \vdots & \ddots & \vdots \\ 0.36 & \cdots & 0.09 \end{pmatrix}$
Hypothesis 2
Estimation Score

Exceptional Sequential Link Trails

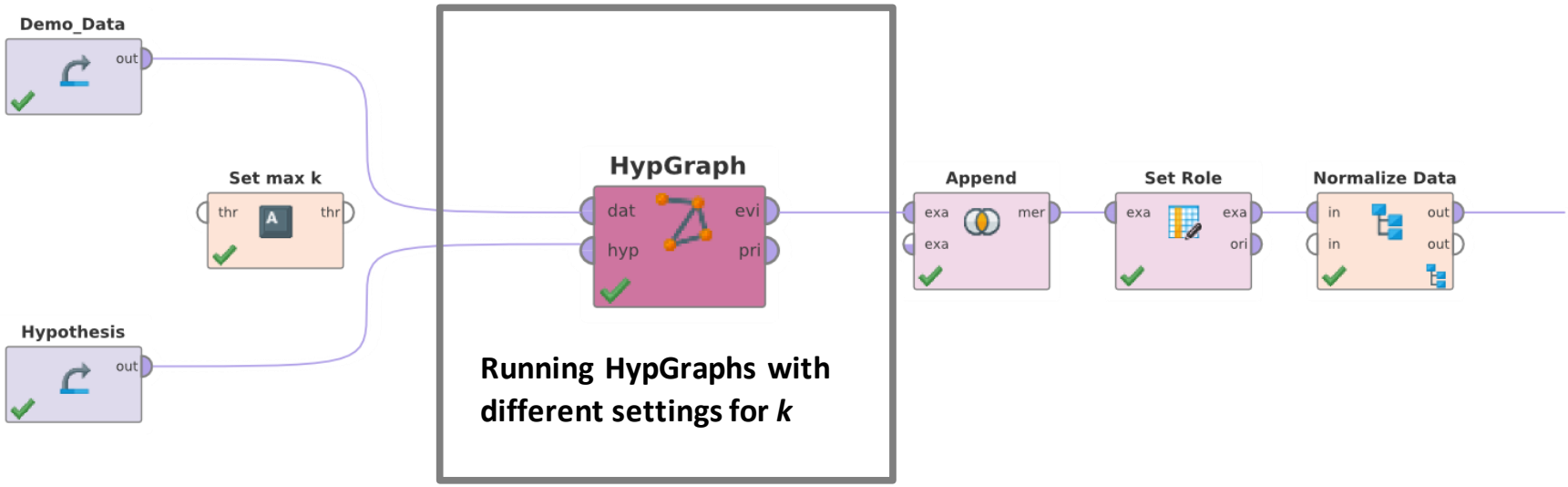
- Construction of the transition matrix:
  - Graph constructed from alarm sequences
  - Considering *subplant – subplant* relations extracted from the P&IDs

- Anomaly detection/analysis:
  - Situations can be evaluated using Bayes factors analysis
  - Comparing data to model hypothesis (normal behavior) and a random one (lower bound)

Exceptional Sequential Link Trails (2)

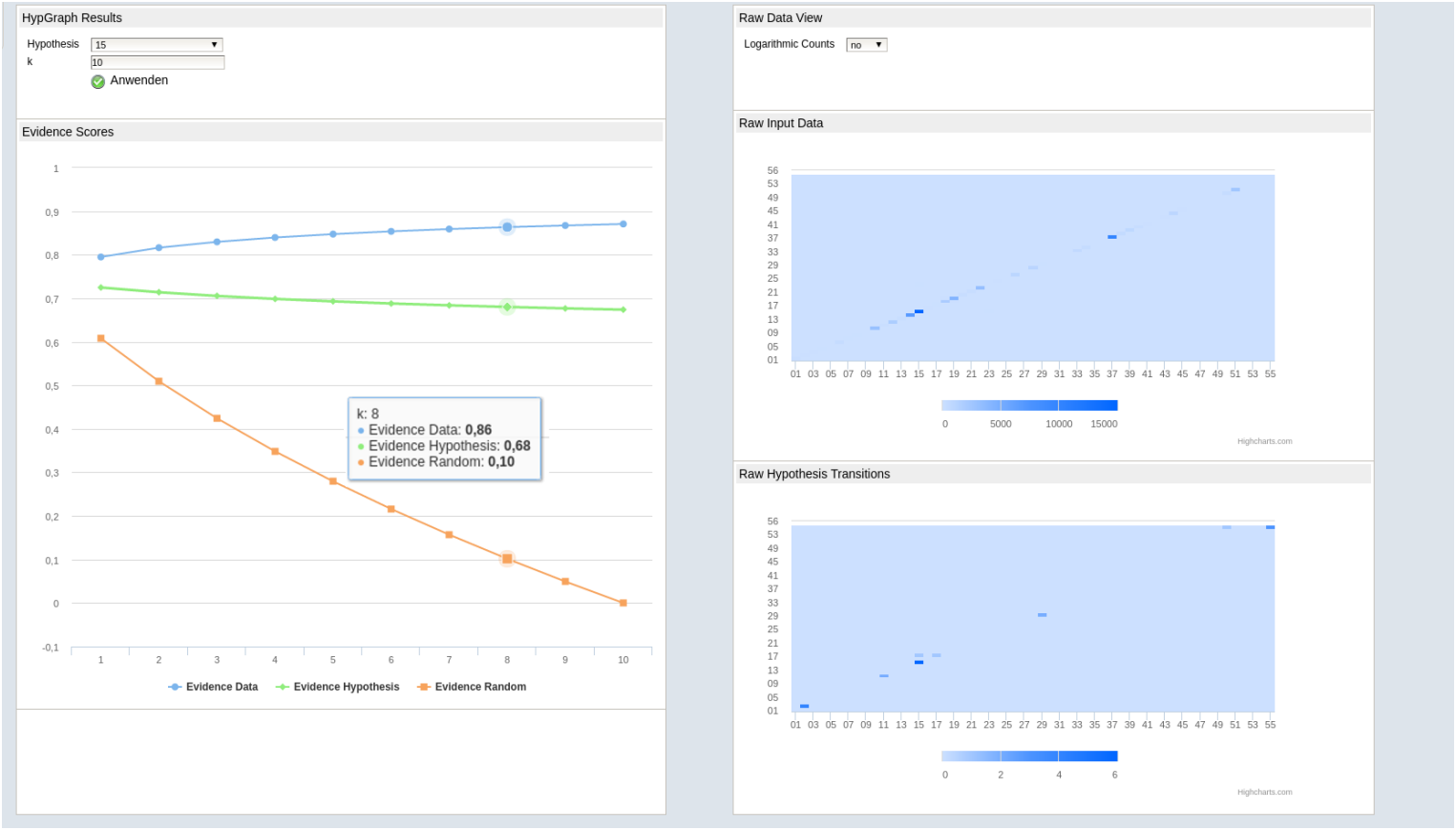


RapidMiner Workflow for HypGraphs



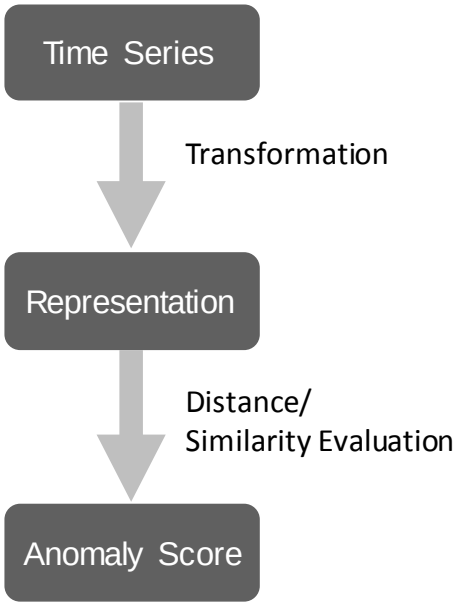
- Varying tolerance factor ***k*** of the estimation algorithm
- Algorithm implemented as RapidMiner extension, available at: **GitHub**

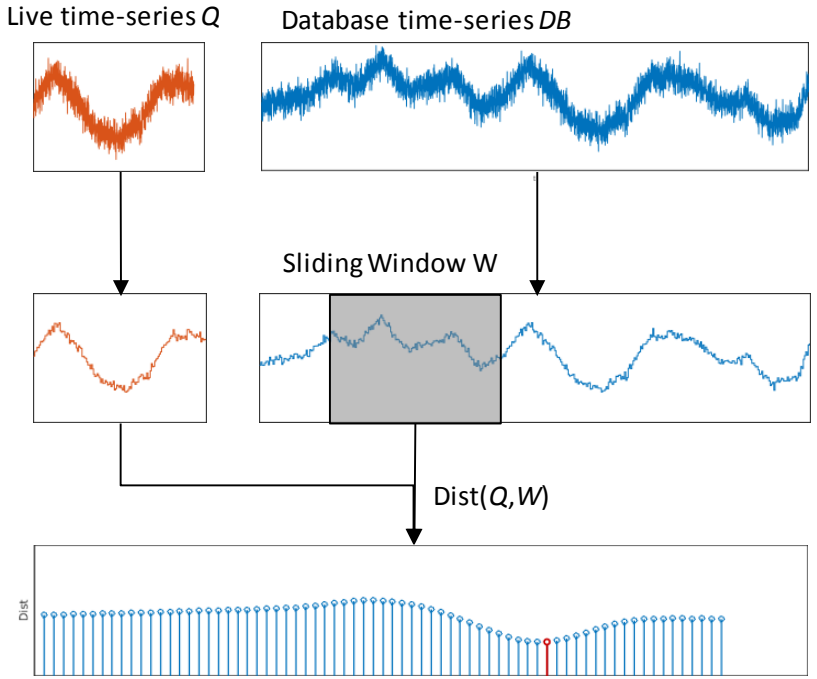
HypGraphs - Visualization



Anomaly Detection on Sensor Data

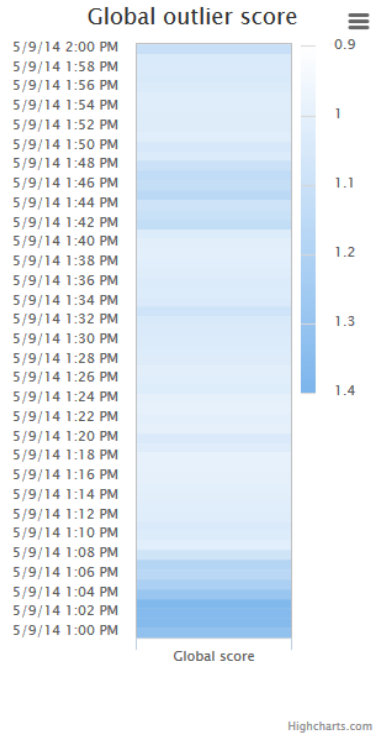
The distance between a live data time-series and the most similar subsequence from historical database is used to calculate the anomaly score.



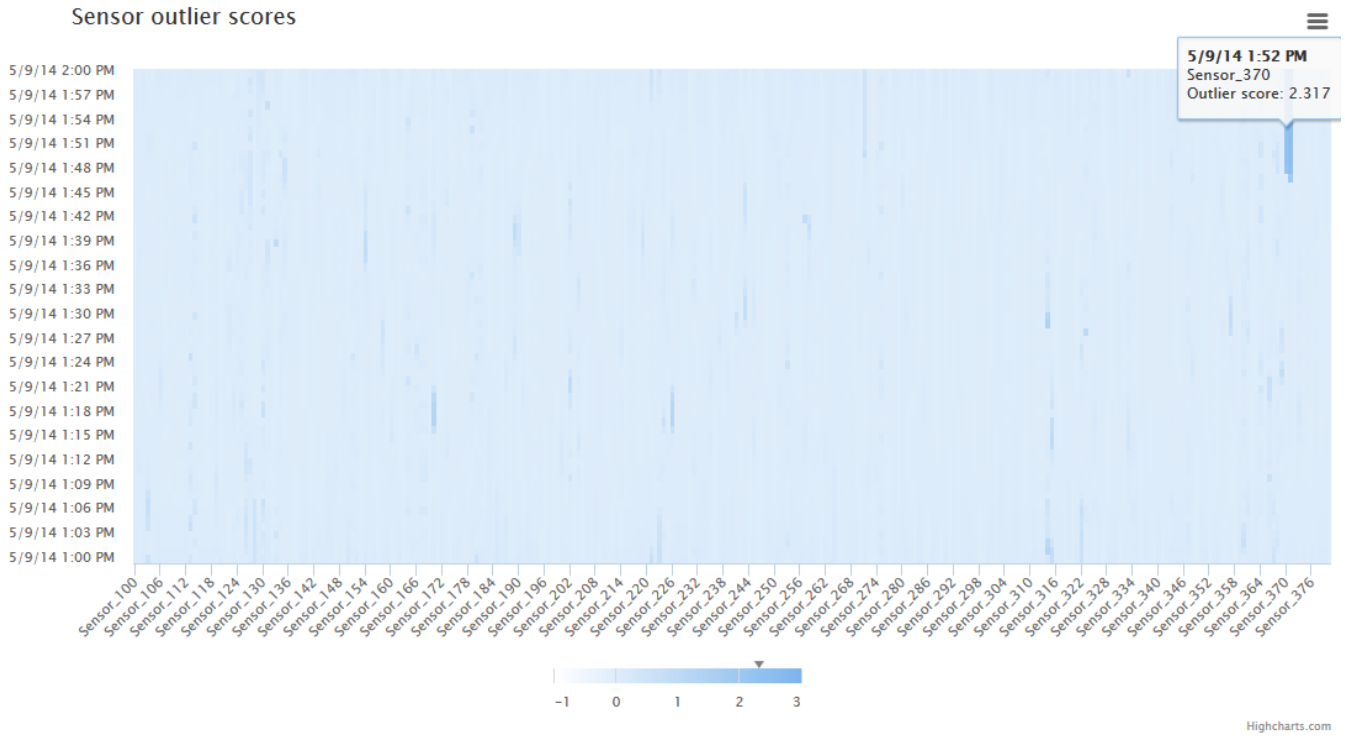


Uni- and Multivariate Anomaly Scores

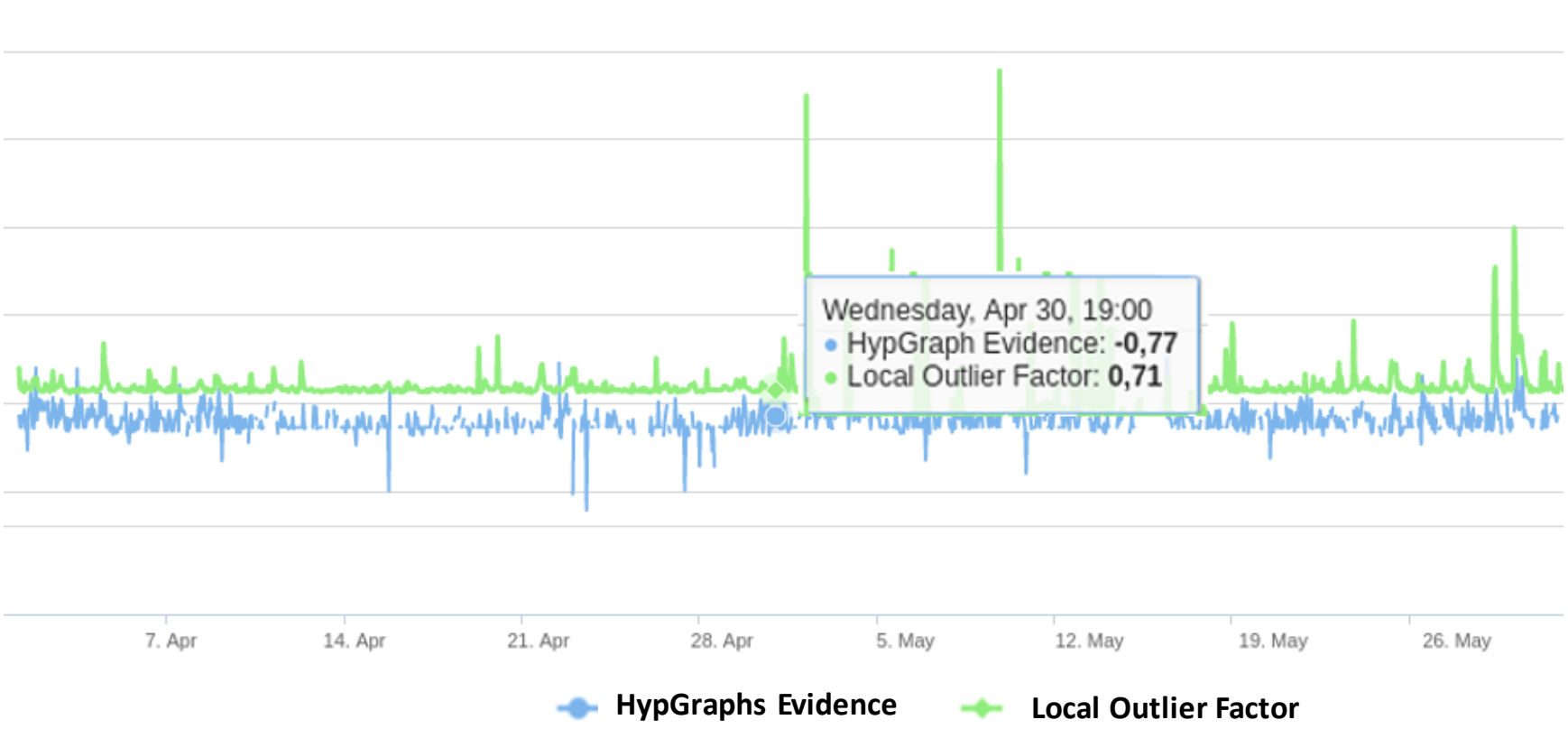
Global outlier score



Single sensor outlier scores



Comparing HypGraphs and Local-Outlier-Factor



Conclusion

- Two approaches for anomaly detection in industrial enviroments
  - HypGraphs: new method for analysing sequential & graph based data
  - Anomaly scores of sensor data

- References:
  - Atzmüller, M. ; Schmidt, A. ; Klöpper, B. ; Arnu, D.:
    HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses.
    In: New Frontiers in Mining Complex Patterns, Postproceedings NFMCP 2016
  - *RapidMiner HypGraphs extension*: https://github.com/rapidminer/rapidminer-extension-hypgraphs

ANOMALY DETECTION AND STRUCTURAL ANALYSIS IN INDUSTRIAL PRODUCTION ENVIRONMENTS

rapidminer
www.rapidminer.com

David Arnu

Email: darnu@rapidminer.com
Web: rapidminer.com

RapidMiner GmbH

Westfalendamm 87
44141 Dortmund
Germany

+49 231 292 993 01

RapidMiner, Inc.

10 Fawcett St., 5th Floor
Cambridge MA 02138
United States

+1 617 401 7708