Anomaly Detection and Predictive Maintenance

Rosaria Silipo
Iris Adae
Christian Dietz
Phil Winters

Rosaria.Silipo@knime.com
Iris.Adae@uni-konstanz.de
Christian.Dietz@uni-konstanz.de
Phil.Winters@knime.com

KNIME.com AG, Zurich, Switzerland
www.knime.com
@KNIME
Text Mining Meets Network Mining

Text Mining for Sentiment + Network Mining for Relevance =
KNIME UGM 2013

Time Series + Machine Learning + Big Data

• Rosaria Silipo + Phil Winters
  – Manufacturing
  – Chemical
  – Life Science
  – Transportation
  – Utilities
  – Automotive
  – Cyber Security

• The Irish Energy Trials
The Internet of Things

Rosaria Silipo
Aaron Hart
Phil Winters
White Papers, Public Data and the KNIME Workflows!

White Papers

Useful white papers from KNIME.

2015

KNIME opens the Doors to Big Data. A practical Example of Integrating any Big Data Platform into KNIME

2014

Taming the Internet of Things with KNIME

Geolocation of KNIME Downloads as a static Report and as a Movie

2013

Big data, Smart Energy, and Predictive Analytics

Analyzing the Web from Start to Finish: Knowledge Extraction from a Web Forum using KNIME

2012

Usable Customer Intelligence from Social Media Data: Clustering the Social Community

Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining

The KNIME Text Processing Feature: An Introduction

2011

Social Media, Recommendation Engines and Real-Time Model Execution with KNIME and ADAPA
KNIME UGM 2015: The Hot Topics....

- Intrusion Detection / Prevention
- Fraud Detection / Prevention
- Fault Detection / Prevention
- Failure Detection / Prevention
- Health Monitoring / Early Warning
- Eco-System Disturbances / Prediction
- Preventive Maintenance
- Predictive Maintenance
- Reliability Analysis
- Anticipatory Failure Determination
- Guided Error Prevention
"The ideal prevention would predict all failure before it occurs"
Anomaly Detection: The Approaches

1. **Supervised Anomaly Detection.**
   A group of patterns are labelled as anomalies and we need to find them. This is just a classification problem where one of the classes is named "anomaly".

2. **Static Unsupervised Anomaly Detection.**
   There are a number of labelled pattern classes and suddenly a weird unrecognized outlier pattern shape shows up. Like an unknown heartbeat during an ECG. Either benign or worrisome, an alarm must be triggered.
3. Dynamic Unsupervised Anomaly Detection.

Here some measures change over time till their values are not normal anymore. For example, while a motor is slowly deteriorating, one of the measurements might change till it gets out of control and the motor breaks. We want to stop the motor before it completely breaks producing even more damages.

This problem is similar to number 2 but slightly more challenging because it is not pattern based and it changes slowly over time.
Dynamic Unsupervised Anomaly Detection Applications
The Fourier Transform

Every signal can be filtered into a series of circular paths:

• How Big? (amplitude)
• How Fast? (Frequency)
• Where to Start (Phase Angle)
The Real Challenge…….

NO PUBLIC DATA !!!
Until Now! A Motor and its Sensors

28 time series from 28 sensors on 8 different parts of a mechanical engine.

A1 (input shaft vertical)
A2 (second shaft horizontal upper bearing)
A3 (third shaft horizontal lower bearing)
A4 (internal gear 275 degrees)
A5 (internal gear 190.5 degree)
A6 (input shaft bearing 150)
A7 (input shaft bearing 151)
M1 (torque KnM)
The Data

• Time Series are FFT-derived Spectral Amplitudes
• There is only one motor breakdown episodes on July 21, 2008
• The breakdown is visible only from some sensors and only in some frequency bands
• The engine was substituted with a new one after the breakdown
Align Time Series
Combine all Spectral Time Series

- Data Cleaning and DateTime Conversion
- Frequency Binning
- Average Spectral Amplitude on Frequency Bins vs. Date
Pivoting

Average Spectral Amplitude by Day and Frequency Bin

- Time: 01.01.2007 to 02.01.2007
- Frequency Bins: [0-100], [100-200], [200-300], ..., [1100-1200], [> 1200] Hz
Time Alignment: Column Merger Node

- Outer join of all column values
- Union of datetime values in one single column
Data Visualization
Data Visualization: Time Plots by Frequency Bands

Only some Spectral Time Series shows the break down
Heatmap of spectrum amplitudes by frequency bands for A1 – SV3 sensor signal

Frequency bands [200,300] Hz and [500,600] Hz show the breaking point in red
Scatter Matrix

- **Jan.01.2007 – Aug.31.2007**
- **Sep.01.2008 – Jul.21.2008**
- **Jul.22.2008 – Apr.20.2009**
Data Visualization: Auto-Correlation Map

A1-SV3 [300-400]Hz signal

A1-SV3 [300-400]Hz signal

Correlation with past values changes in Time!
Data Visualization: Correlation Map

Correlation across Frequency Bands changes in Time!
Data Analytics
The Approach

• Read all data
• Spectral amplitudes on frequency bins vs. time
• Define “normal”
• Learn “normal” values and patterns
• Detect alarming situations
• Export models as PMML
Learn “normal” to recognize what is not

- We only have normal measures/patterns available till the system starts failing
- We need to train a machine learning algorithm to predict/classify those “normal” patterns/measures
- We then use the machine learning algorithm’s uncertainty to fire an alarm on what might not be “normal”
Learn “normal”: Training Set

Only some Spectral Time Series shows the break down
Time Series Prediction
Learn “normal”: Input and Target Features
Time Series Training: Learn what is “normal”

On each Frequency Band, on each Sensor Signal, on each Motor Part:

Train a Linear Auto-Regressive Model

(Lag = 10, no seasonality)

595 AR Models
Time Series Training: Learn what is “normal”

- 10 past values
- Last value for missing values
- Train 595 AR Models And Write to PMML files
Time Series Production: Linear Auto-Regression Prediction

- **Column List Loop Start**
  - Loop on all columns i.e. on all frequency bins for each sensor
  - [500-500]-Amp -> Amp

- **Lag Column**
  - Lag = 10

- **Missing Value**
  - Time Series
  - Jan-Aug 2007 training set
  - Extract Time Window
  - Linear Regression
  - Node 178
  - Linear AR(p) p=10
  - and first level alarms from prediction errors
  - abs and squared prediction error

- **Extract Time Window**
  - Sep 2007 - Apr 2009 remaining data

- **Java Edit Variable**
  - Column Rename Loop End (Column Append)
  - back to original column name
  - Node 244

- **Prediction, Prediction errors, First level alarms**

- **Line Plot**
  - A7, SA1 [200,300]Hz
  - Apr 15 2008
  - Original signal
  - Jul 21 2008
  - Prediction
  - Sep 2007
  - Apr 2009
KNIME in Action

Learning/Testing

Time Series

Read Data → Align Times → Visualize

Perform FFT if necessary

Across all Sensors

join models

The Signal Model

Table Writer

The Historical Data

PMML Writer

PMML Reader

The Signal Model

Production Process

Node 37

Table Reader

The Historical Data

Signal Execution

Node 40

PMML

Production
Signal Execution: Part 1
Time Series Production: Recognize what is not “normal”

On each Frequency Band, Sensor Motor part:

Prediction Error PE(t) as:
\[ PE(t) = \text{abs}(\text{Prediction}(t) - x(t)) \]

First Level Alarm
IF \( PE(t) > \text{Mean}(PE(t)) + K \times \text{stddev}(MA(t)) \) => \( alm(t) = PE(t) \)
ELSE \( alm(t) = 0 \)

\[ MA(t) = \text{Moving Average}(alm(t), N=21, \text{backward window}) \]

Second Level Alarm
IF \( MA(t) > 0.01 \) => Alarm
Time Series Production: Second Level Alarms

If alarm => send email

- Sep 2007 - Jul 2008 Predictive Maintenance Time Window
- Column Filter: only first level alarms
- Moving Aggregation: moving average backward window N = 21
- Column Filter: only Mean(alarm)
- Second Level Alarms: detect second level alarm and send email

- A7, SA1 [200,300]Hz
- A7, SA1 [400,500]Hz
- A7, SA1 [300,400]Hz

- May 05 2008
- Mar 06 2008
- Jul 2008

Sep 2007
Time Series Production: Second Level Alarms R Stacked Plot

Second Level Alarms. Stacked Plot.

MA(\text{alm}(t))

May 05 2008

Mar 06 2008

A7, SA1 [200,300]Hz
Signal Execution: Part 2
Time Series Production: Use KNIME to Send Alarms !!!

Using RESTful Services:
But we can go Further!  

Pattern Recognition

Time series prediction

Input

target

Input/target

Pattern Recognition

spectral amplitude

Scale
Pattern Recognition

Iris Adae
Recognize what is not “normal”

Pattern Recognition

• Output Probability / Confidence < threshold => Alarm

• Auto-associator pattern reproduction error > threshold => Alarm
Pattern Recognition: General Windowing Scheme

- Model the system with data from the past
- Use the system with most recent data point


Single observation on multiple frequency bands
• Fuzzy c-means -> clusters
• Fuzzy c-means also generates an outlier cluster
• If recent data point is in outlier cluster -> Anomaly
• Results: Fuzzy Cluster
Pattern Recognition: Similarity Based

- Similarity Search -> nearest neighbor to current pattern
- If no nearest neighbor -> Anomaly
Pattern Recognition: Similarity Based

Apr 14 2008

Jul 21 2008
KNIME in Action
But KNIME can do even more !!!

Slow motion cameras can now record \(4,400,000,000,000\) frames per Second

![Image of a mug with "10,000fps"]

We don’t have images of the motor turning...... but what if we did ?
Image Processing

Christian Dietz
The Dataset

Data from DAGM 2007 Competition
(data sponsored by Robert Bosch Corporate Research department, Schwieberdingen, Germany)

Goal: Find defects in images with minimal user-interaction

Dataset contains about 1100 images.
What’s a Defect?
How to detect them in KNIME?
Reading Images
Calculating Features

For each pixel we calculate features in the neighborhood of the pixel.

**Examples:** Intensity Statistics (Mean, Variance, StdDev, etc), Texture, ...

**Computational expensive step:** We are working on a faster version.
Annotate Training Data

- Read in Images
- Fiji Trainable Segmentation Features 2D
- Annotate and Split into Training & Testing
- Node 27
- Node 36

Defect
Not a defect
Given the training data we can train a model to distinguish defect pixels from not defect pixels.

**Model:** Random Forest

**Additional:** We determine the average size of the defects for pruning too small and too big defects
Detecting Defects
Detecting Defects: pass the new column back to learning!
KNIME in Action

Diagram showing processes and nodes for data handling and analysis.
Possible Next Steps.....

Random Forrest and Threshold Detection
Combining the Models
Text Processing
Big Data
Thank You

We will announce White Paper availability!

Questions:
• [http://tech.knime.org/forum](http://tech.knime.org/forum)
• Rosaria.Silipo@knime.com
• Phil.Winters@knime.com

Follow us:
• Twitter: @KNIME
• LinkedIn: [https://www.linkedin.com/groups?gid=2212172](https://www.linkedin.com/groups?gid=2212172)
• KNIME Blog: [http://www.knime.org/blog](http://www.knime.org/blog)