Sensor time series predictive analysis

MTA **SZTAKI**
Hungarian Academy of Sciences
Institute for Computer Science and Control

András Benczúr, head, Informatics Lab

Anna Mándli, Ph.D. student, Bosch

Péter Vaderna, Ericsson

Bálint Daróczy, postdoctoral researcher, computer vision, deep learning

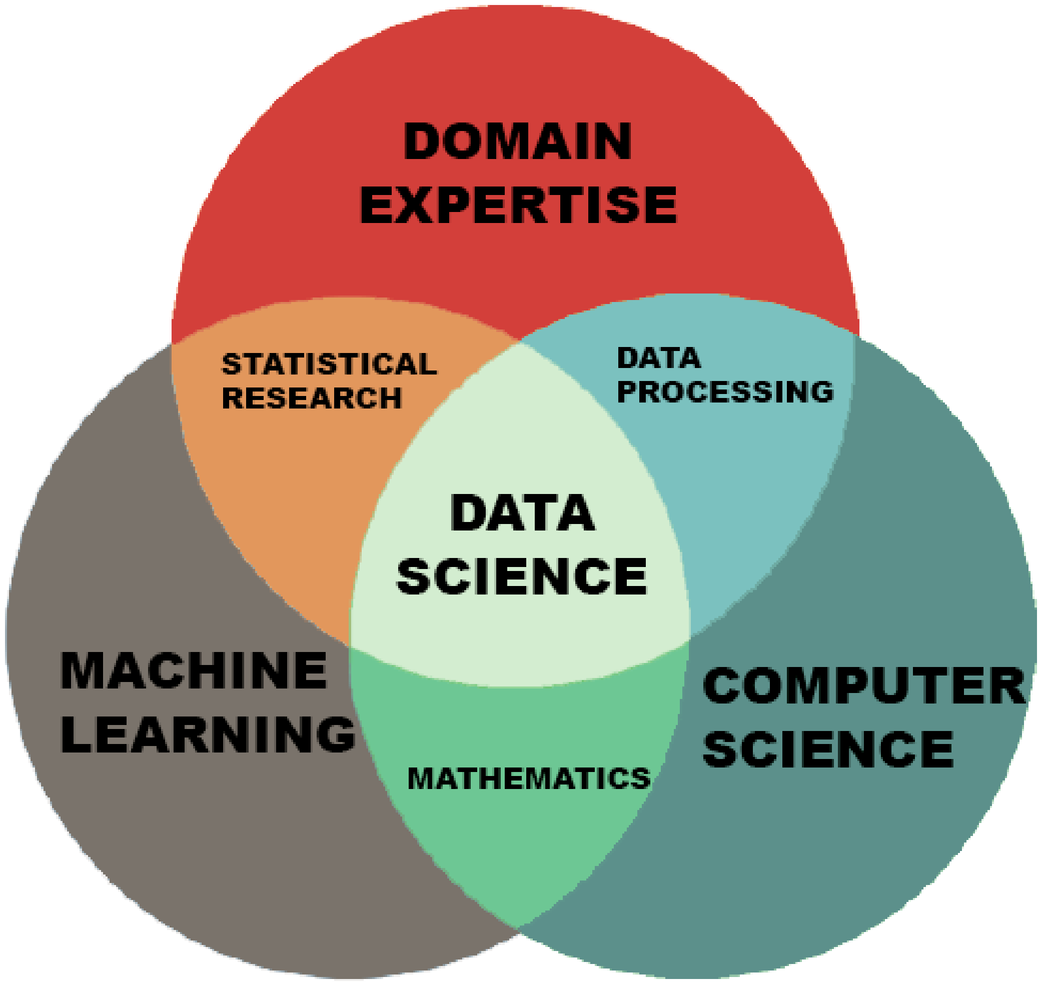
Róbert Pálovics, postdoctoral researcher, recommenders, online machine learning

October 2, 2017

Overview of the Lab

Applied data science lab

- Machine learning
  - Time series
  - Image processing
  - Kernel methods
- Computer science
  - Scalable software systems (Hadoop, Spark, Flink)
- Domain expertise
  - Manufacturing
  - Sensors, IoT
  - Vision



Main partners

- Industry in Hungary
  - Bosch: scrap rate prediction and root cause analysis
  - OTP Bank: modeling by machine learning
  - AEGON: customer data fusion, fraud detection
  - Telecom, Vodafone: search engine
  - Ericsson
    - Mobile session drop prediction
    - Data Streaming access and analytics
- International
  - EIT Digital and H2020 Big Data activities
  - TU Berlin, Portugal Telecom, Rovio, Atos

Contents

- Ericsson use case: mobile radio session drop – Introduction
- Methods for time series classification
  - Traditional methods
  - Support Vector Machines
  - The similarity kernel
- Radio Session drop prediction experiments
  - Ericsson: based on cell tower data
  - Samsung: based on smartphone logs
- Bosch use case: scrap rate prediction in transfer molding

MTA SZTAKI
Hungarian Academy of Sciences
Institute for Computer Science and Control

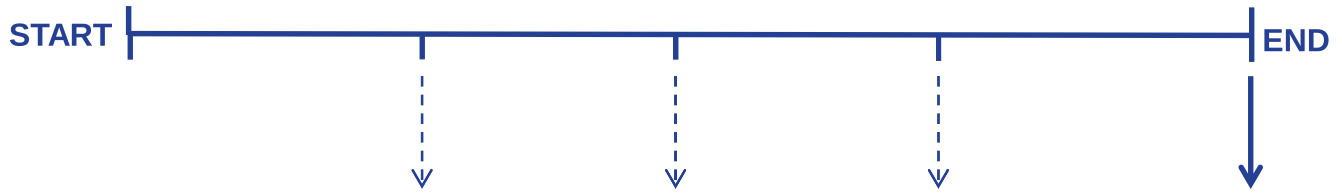
Use Case 1:
Mobile radio session drop

ERICSSON

Data is based on eNodeB CELLTRACE logs from a live LTE (4G) network

RRC connection setup /
Successful handover into the cell

UE context release/
Successful handover out of the cell



Per UE traffic report
(traffic volumes, protocol events (HARQ, RLC))
Per radio UE measurement
(CQI, SINR)
Period: 1.28s

Per UE measurement reports
(RSRP, neighbor cell RSRP list)
Configurable period
Not available in this analysis

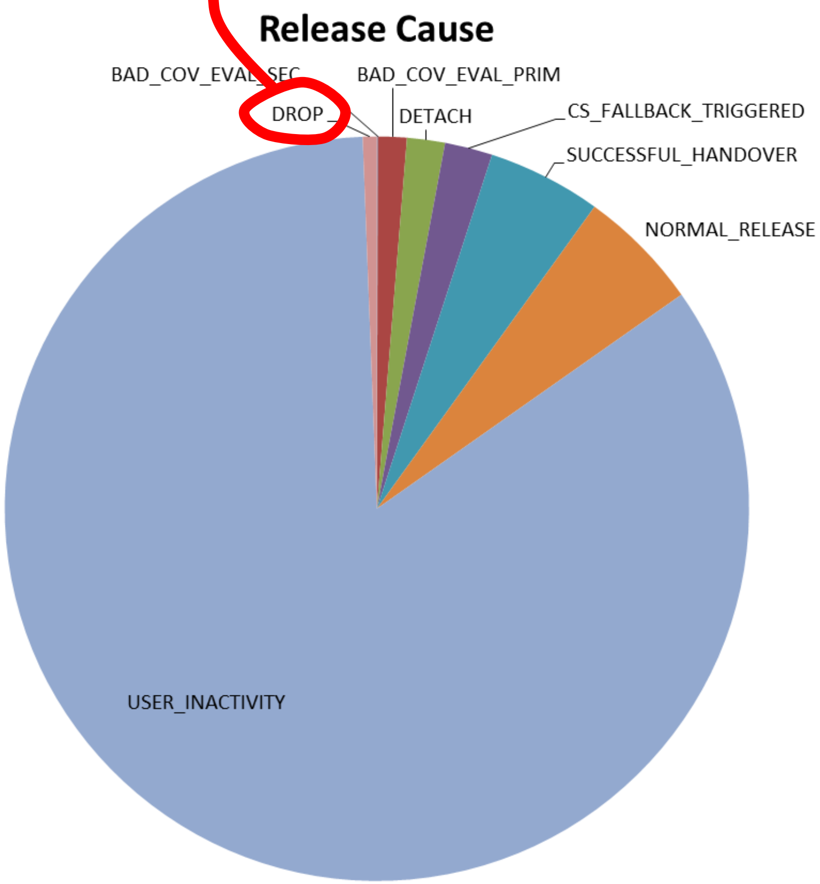
Session record example

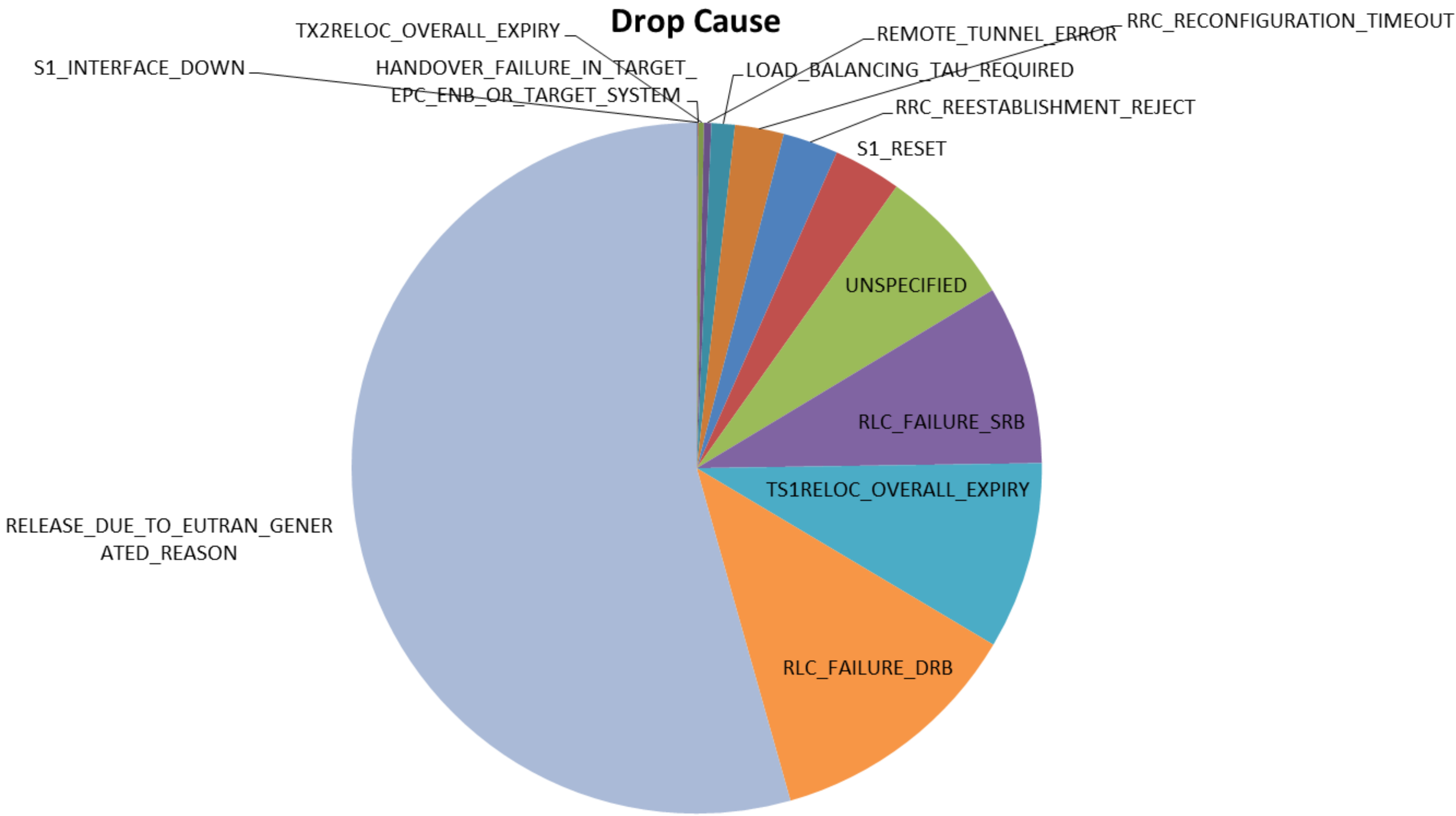
	Variable	Range	Comment
Time & Location	Timestamp (s) Duration (s) User Id Cell Id	 0.1 - 30	UNIX timestamp
Dependent variables	release cause drop category	~20 different cause codes 0 (no drop), 1 (IRAT change), 2 (drop)	
Independent variables*	cqi_avg harqnack_dl harqnack_ul rlc_dl rlc_ul sinr_pusch_avg (dB) sinr_pucch_avg (dB)	1 - 15 0 - 1 0 - 1 0 - 1 0 - 1 -4 - 18 -13 - 3	Channel quality index HARQ NACK ratio in downlink HARQ NACK ratio in uplink RLC NACK ratio in downlink RLC NACK ratio in uplink Signal to Interference plus Noise Ratio on Uplink Shared Channel Signal to Interference plus Noise Ratio on Uplink Control Channel

Independent variables are time series (arrays) of varying length

Root-cause analysis: Cause of Release and Drop

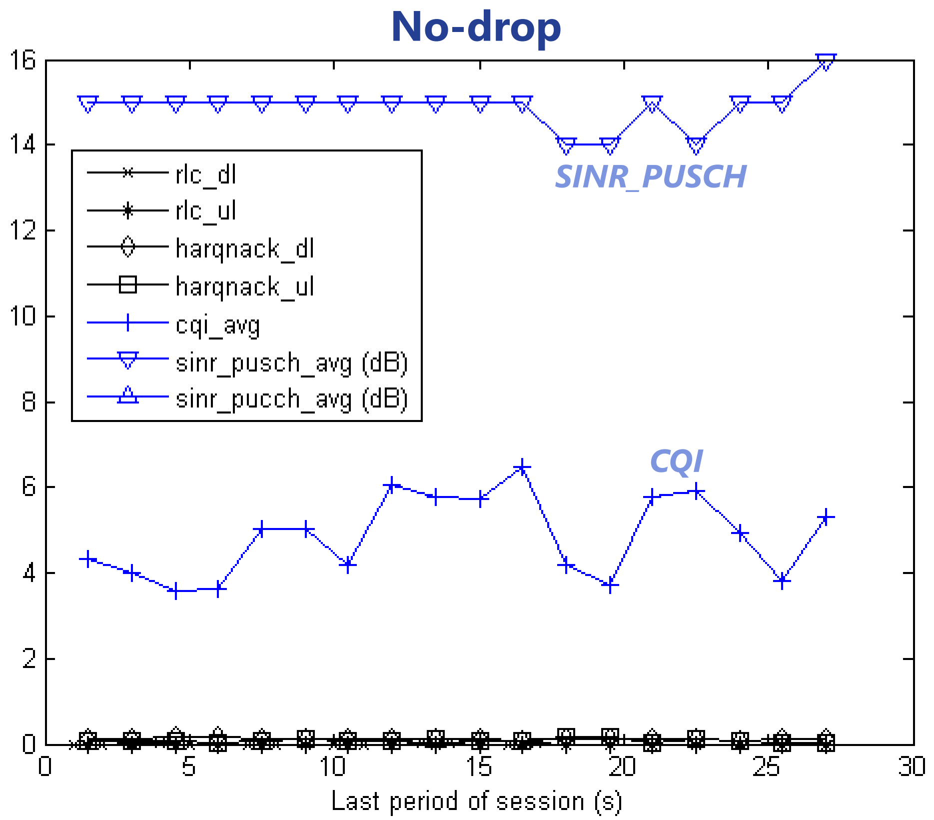
0.6% of sessions are dropped

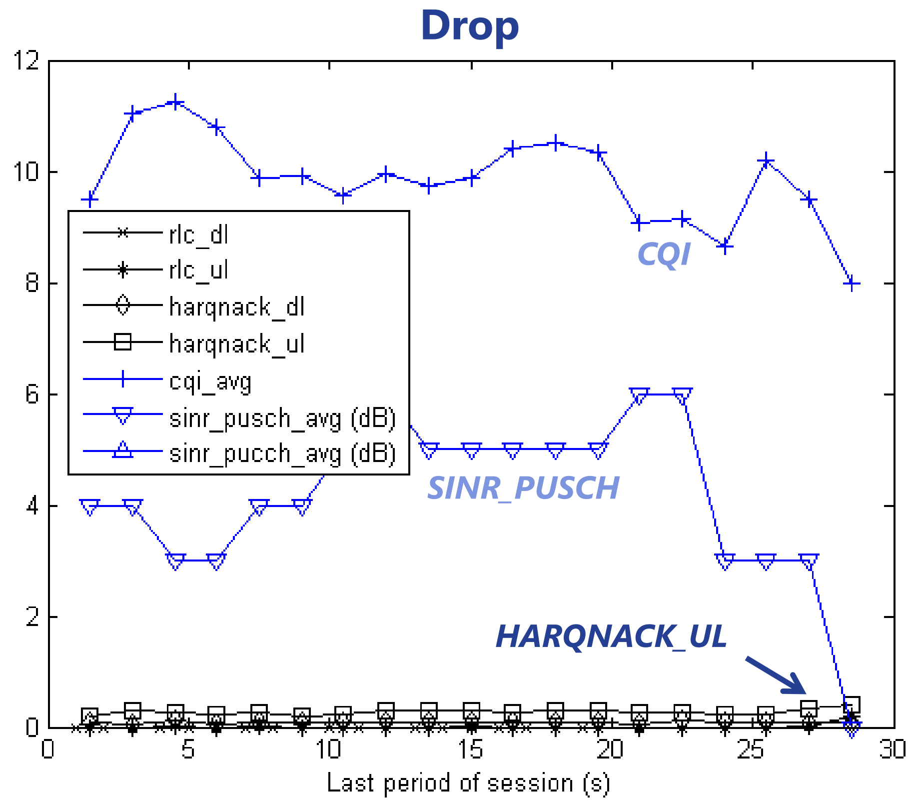




Drop prediction

- Detailed dataset (report every 1.28 sec)
- Examples for typical individual cases (last 30 sec)





Setup for machine learning

Labeled set of sessions (R):

Session 1:
sinr_pusch [s11, s12, …]
rlc_dl [r11, r12, …]
…
drop: yes/no

Session 2:
sinr_pusch [s21, s22, …]
rlc_dl [r21, r22, …]
…
drop: yes/no

⋮

Session R
…

DTW measures

Session to predict:

Session X:
sinr_pusch [sX1, sX2, …]
rlc_dl [rX1, rX2, …]
…

4 algorithms:
Adaboost over statistics (baseline)
Adaboost over Dynamic Time Warping (DTW)
Support Vector Machine (SVM) over statistics
SVM over DTW using kernel method

Class confusion matrix

MTA SZTAKI

		Predicted class	
		Drop	OK
Actual class	Drop	TP: True positive	FN: False negative
	OK	FP: False positive	TN: True negative

- Accuracy: (TP+TN)/(TP+FP+TN+FN)
- Precision (fraction of predicted drop that is actual drop): Prec = TP/(TP+FP)
- Recall (fraction of actual drop identified): Rec = TP/(TP+FN)
- F-measure: Harmonic mean of Precision and Recall
  - F1 = 2 x Prec * Rec / (Prec + Rec)

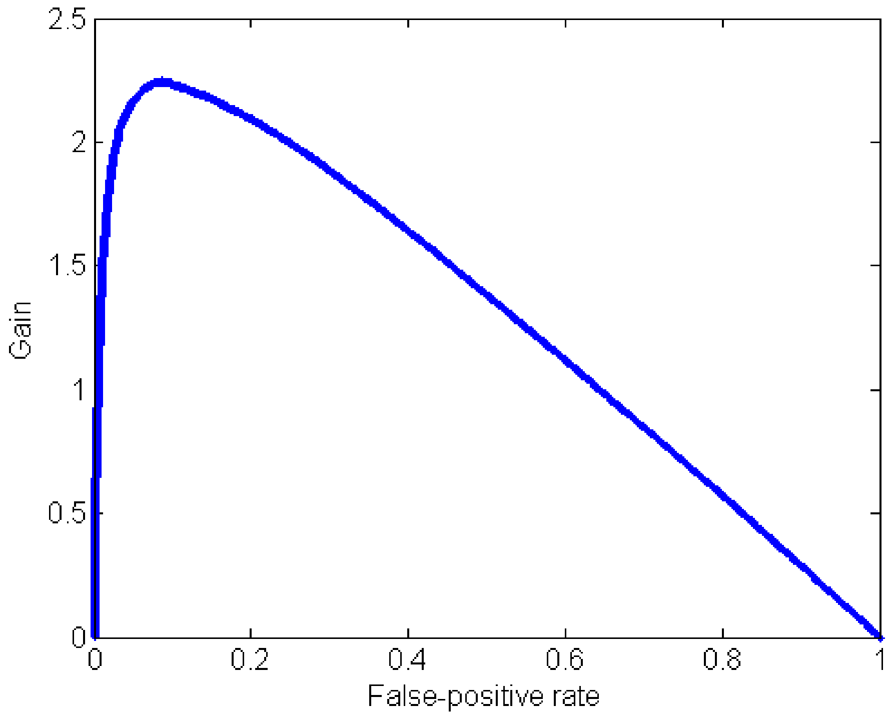
SON aspects

SON aspect: actuation based on prediction (e.g. switching to other channel, assisting in handover decision, etc.)

Gain matrix:

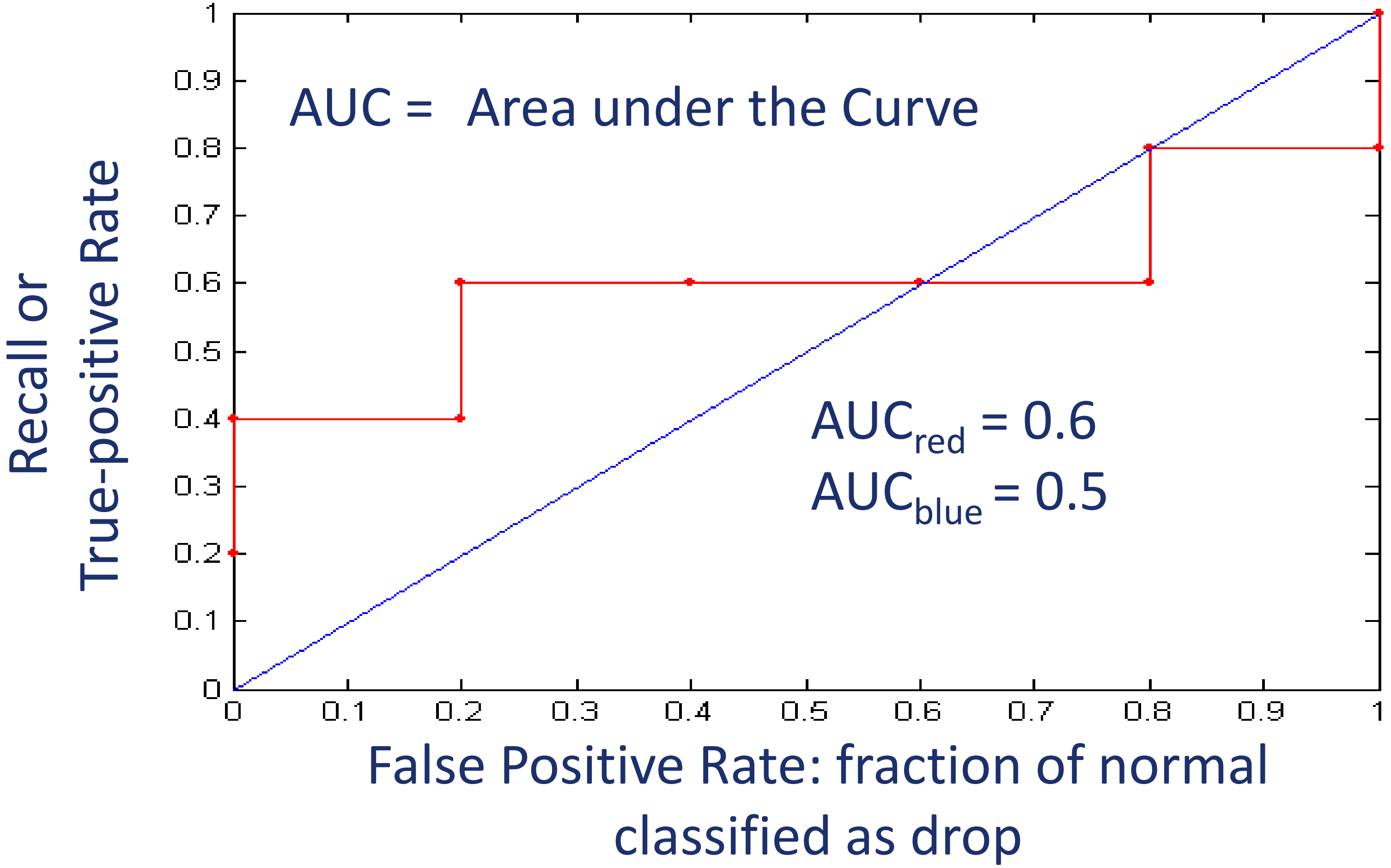
		Predicted class	
		drop	no-drop
Actual class	drop	+2	-2
	no-drop	-2	0

Cost/Gain optimization for SON based on the gain matrix and the ROC curve:



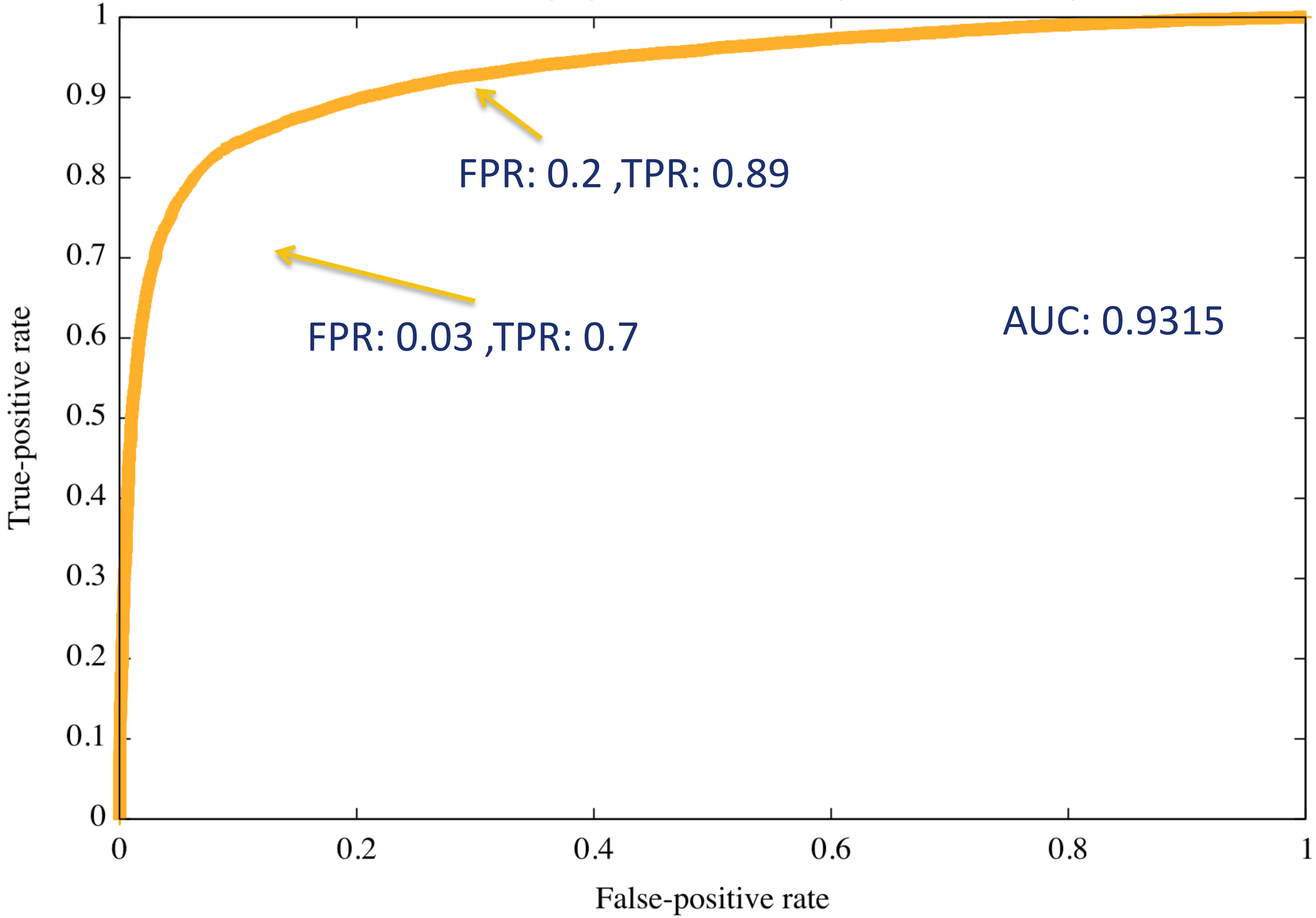
Ranking: Reciever Operating Characteristic

- Based on the ranked order of drop „likelihood”
- No need for decision threshold
- Stable comparison of
  prediction methods



Interpreting AUC for error rates

ROC curve of the best drop predictor 5 periodic reports before drop



AUC ranges:

- 0.5 – 0.55 completely random
- 0.55 – 0.65 impractical
- 0.65 – 0.75 starts working
- 0.75 – 0.85 good
- 0.85 – 0.95 excellent
- 0.95 – 1.0 the task was completely trivial and boring

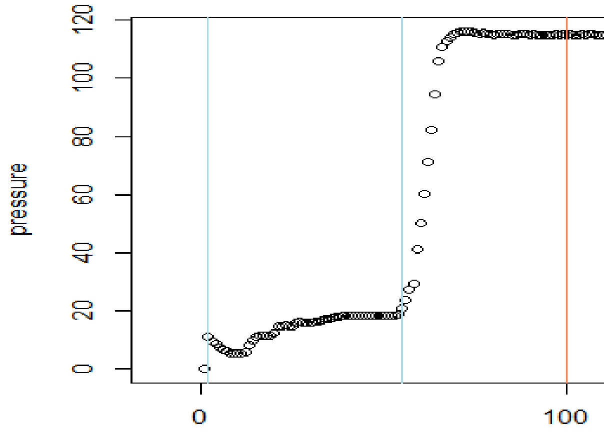
MTA SZTAKI
Hungarian Academy of Sciences
Institute for Computer Science and Control

Time Series classification
A very brief tutorial

Class of methods

Feature based

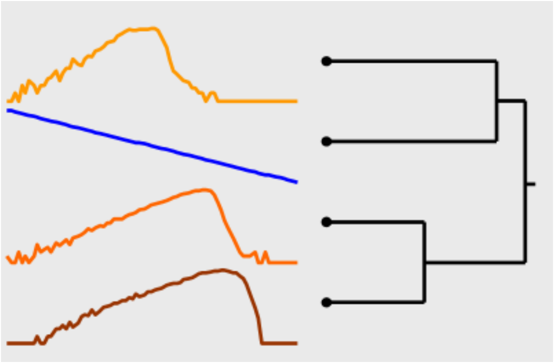
- Minimum, maximum, mean, median, variance, derivatives, …
- Values at characteristic parts of the curve, for example, initial phase in plastic transfer molding, before plunger fully pushes on pellet

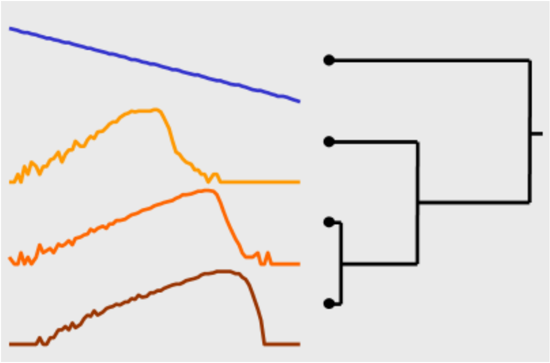


- Numeric or mixed features fit well for decision tree based methods

Shape based

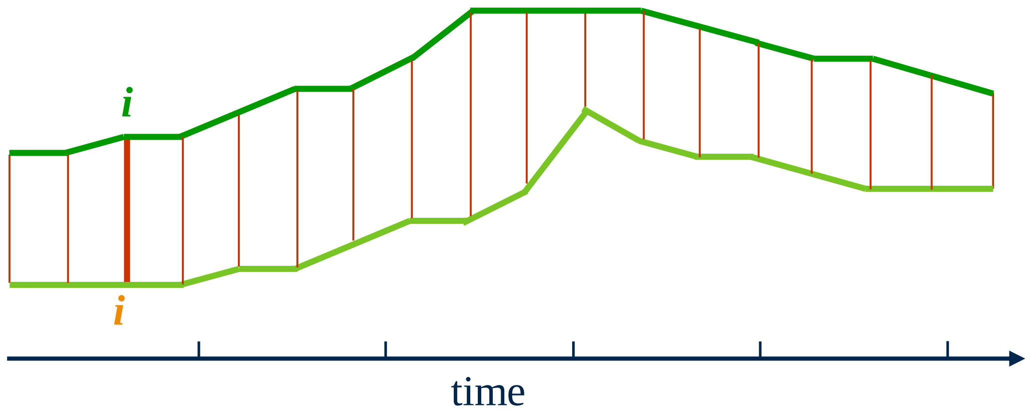
- Subjectivity

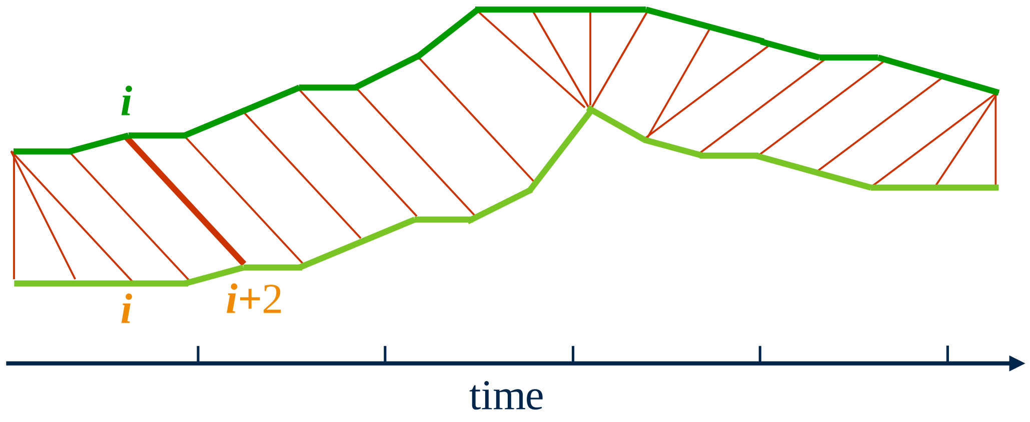




- Several transformations available
  - Offset in value AND in time too
  - Scale, amplitude (unit of measurement)
  - Linear trends
  - …
- Needs good distance measure
- We may use nearest neighbor methods

Dynamic Time Warping: the most prefered distance

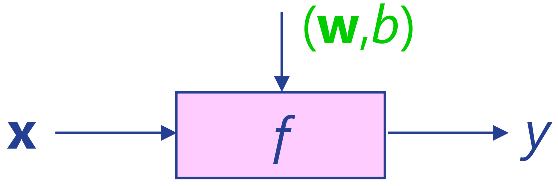




Any distance (Euclidean, Manhattan, …) which aligns the *i*-th point on one time series with the *i*-th point on the other will produce a poor similarity score.

A non-linear (elastic) alignment produces a more intuitive similarity measure, allowing similar shapes to match even if they are out of phase in the time axis.

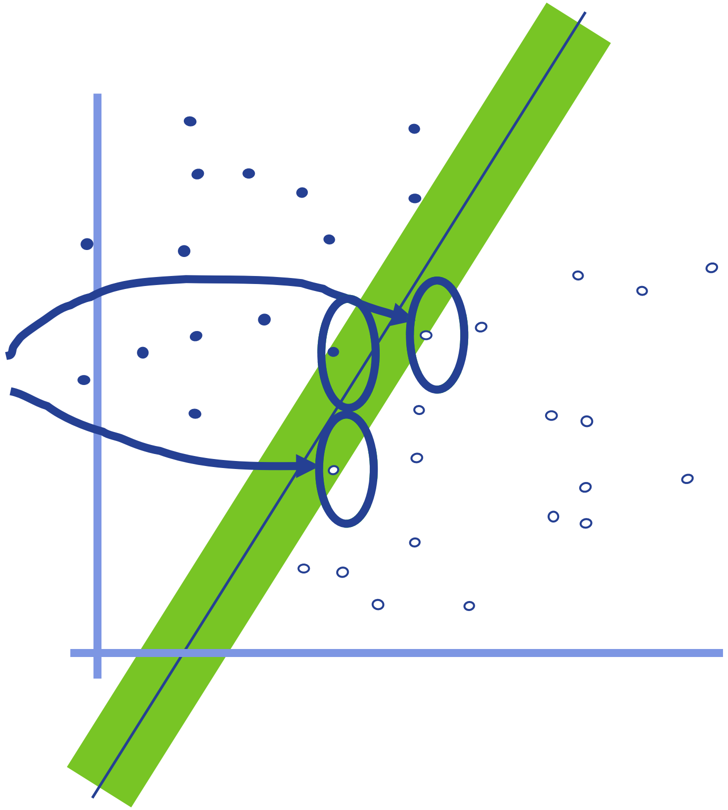
Linear SVM: Maximum margin linear classifier



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}\mathbf{x}+b)$

- • denotes +1
- ° denotes -1

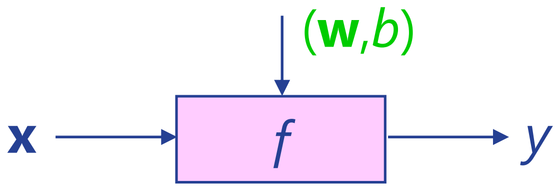
Support Vectors: data points that the margin pushes up against



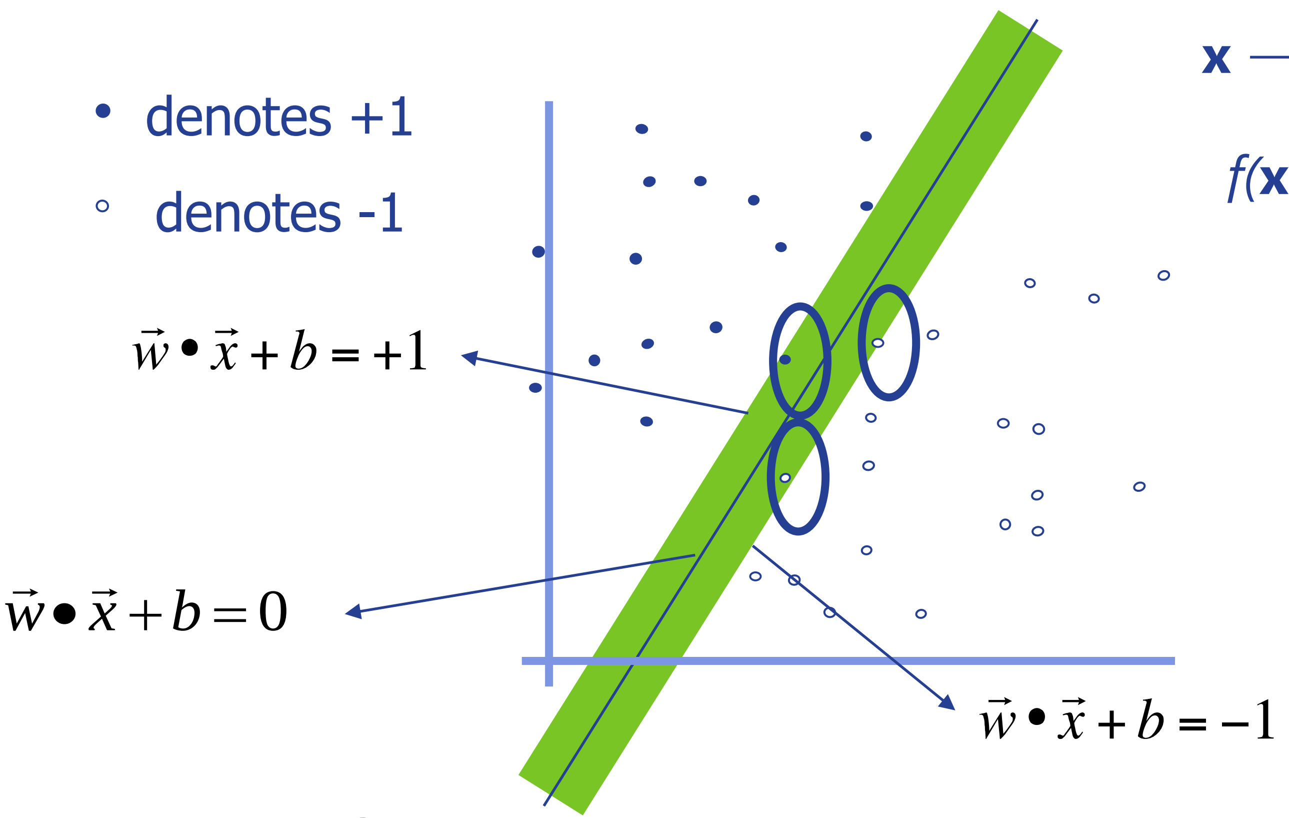
Kernel trick transforms into a high-dimensional nonlinear space

Warmup: linear SVM

- Maximal margin linear classifier



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}\mathbf{x}+b)$



We want to use pairwise similarities to define a kernel space …

Margin can be NEGATIVE!

$$\text{Margin} = \frac{2}{\|\vec{w}\|^2}$$

$$f(\vec{x}_i) = \begin{cases} 1 & \text{if } \vec{w}\bullet\vec{x}_i + b \geq 1 - \xi_i \\ -1 & \text{if } \vec{w}\bullet\vec{x}_i + b \leq -1 + \xi_i \end{cases}$$

Support Vector Machine optimization

MTA SZTAKI

- Need to minimize

$$L(w) = \frac{\|\vec{w}\|^2}{2} + C\left(\sum_{i=1}^{N} \xi_i^2\right)$$

- Subject to

$$f(\vec{x}_i) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x}_i + b \geq 1 - \xi_i \\ -1 & \text{if } \vec{w} \bullet \vec{x}_i + b \leq -1 + \xi_i \end{cases}$$

- But w is a linear combination of the support vectors:

$$\vec{w} = \sum_{i=1}^{N} \alpha_i \vec{x}_i \longrightarrow \|\vec{w}\|^2 = \left(\sum_{i=1}^{N} \alpha_i \vec{x}_i\right)^T \bullet \left(\sum_{i=1}^{N} \alpha_i \vec{x}_i\right) = \sum_{i,j} \alpha_i \alpha_j \vec{x}_i^T \bullet \vec{x}_j$$

- All we need is a way to compute w • x for a support vector w and another vector x → any two data points
- Goal: a mathematically sound notion of w • x based on the similarity of the document!

Support Vector Machine optimization

MTA SZTAKI

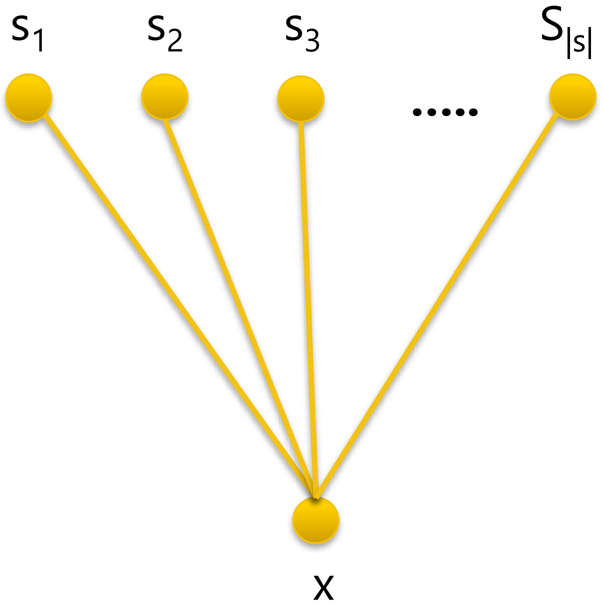
In a transformed space ($x \rightarrow \phi(x)$) we only need the value of the dot product (kernel value):

$$K(x_i,x_j) = \phi(x_i)^T\phi(x_j)$$

$$\vec{w} = \sum_{i=1}^{N} \alpha_i \vec{x}_i \longrightarrow \|\vec{w}\|^2 = \left(\sum_{i=1}^{N} \alpha_i \vec{x}_i\right)^T \bullet \left(\sum_{i=1}^{N} \alpha_i \vec{x}_i\right) = \sum_{i,j} \alpha_i \alpha_j \vec{x}_i^{\,T} \bullet \vec{x}_j$$

Similarity kernel

- Represent session X as Random Field $P(x|\theta)$ generated by distances from other sessions s_1, s_2, …
- Sample set S can be all training set, a sample, or cluster representatives



The simplest Random Field

- Fisher score of X is vector $G_X = \nabla\theta \log P(x|\theta)$
- Fisher matrix is $E[G_X G^T_X]$
- Define a mapping $x \to G_x F^{-1/2}$
- Yields the Kernel $K(x, y) = G_x^T F^{-1} G_y$

Intuitive interpretation: G_X is the direction where the parameter vector θ should be changed to fit best to data point X.

- For a Gibbs Random Field we get a natural normalization of the distances from the sample S:

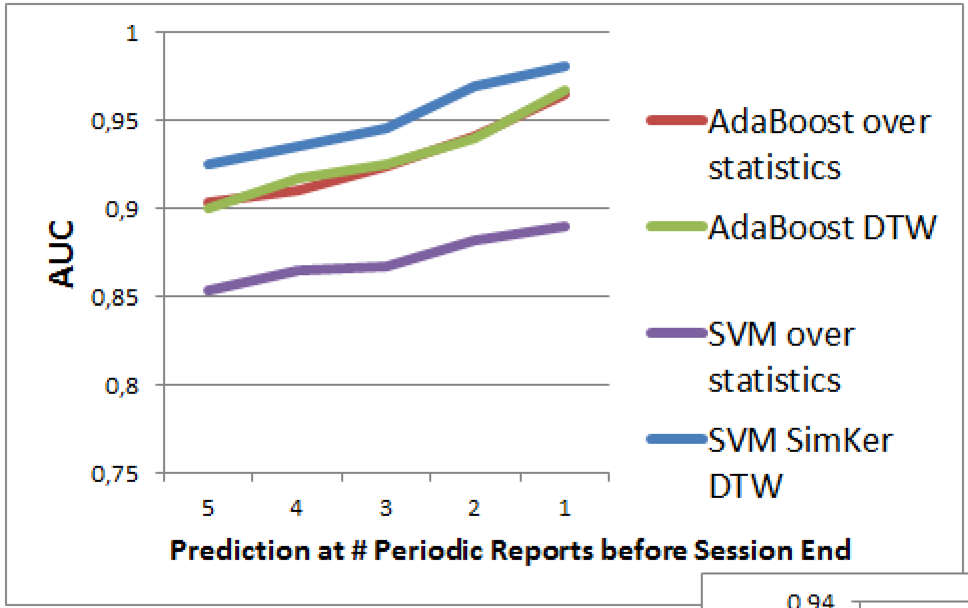
$$\mathcal{G}_X^i = F_{ii}^{-\frac{1}{2}} G_X^i = \frac{\mathbf{E}_\theta[\mathrm{dist}(x, s_i)] - \mathrm{dist}(x, s_i)}{\mathbf{E}_\theta^{\frac{1}{2}}[(\mathbf{E}_\theta[\mathrm{dist}(x, s_i)] - \mathrm{dist}(x, s_i))^2]}$$

MTA SZTAKI
Hungarian Academy of Sciences
Institute for Computer Science and Control

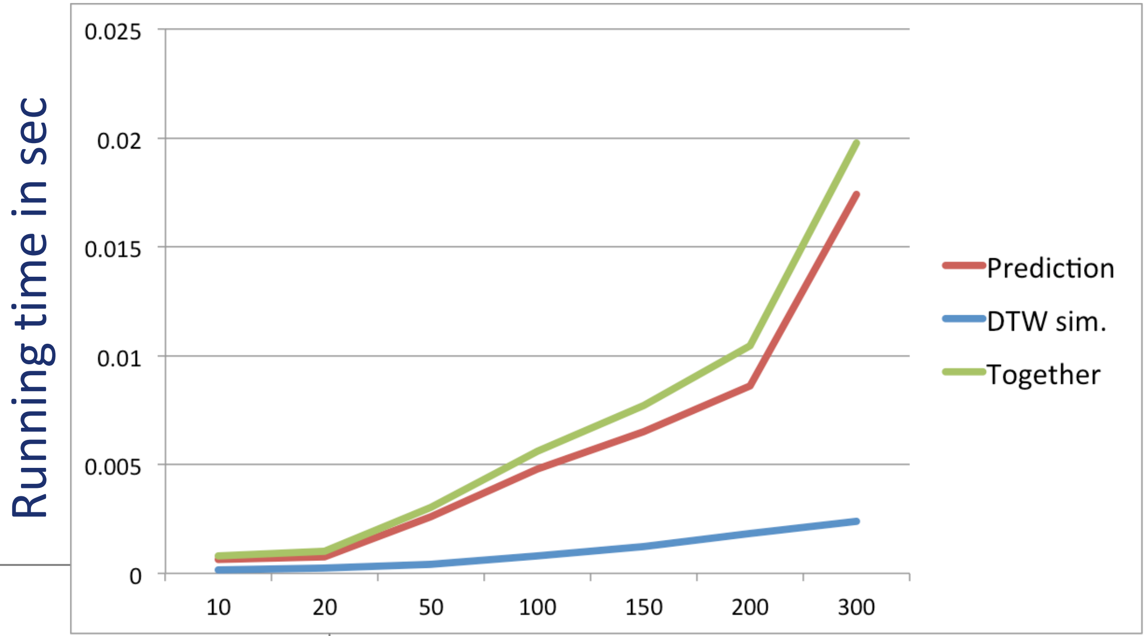
Use Case 1:
Mobile radio session drop

ERICSSON

Evaluation

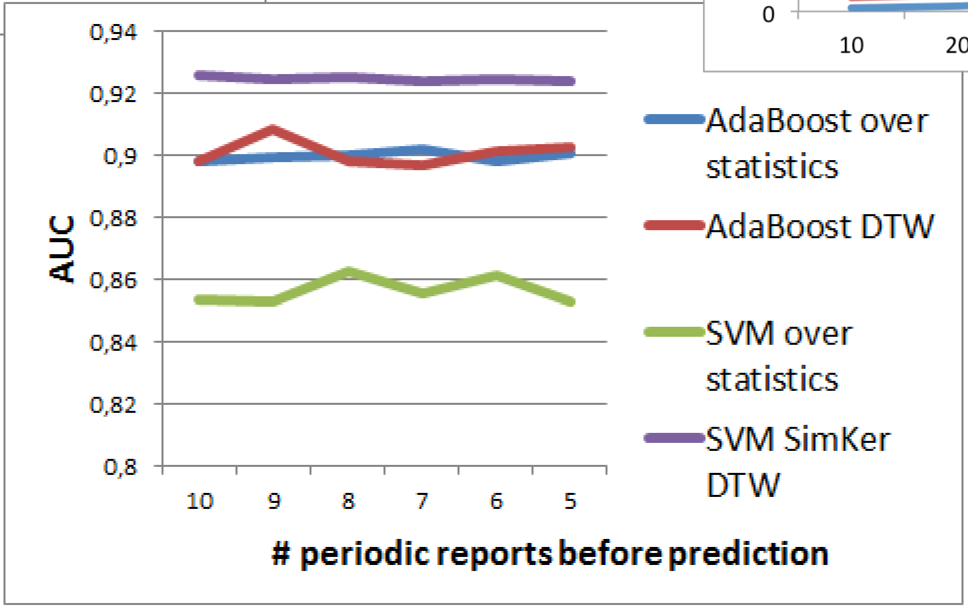


DTW similarity kernel calculations, one session



Sample size |R|

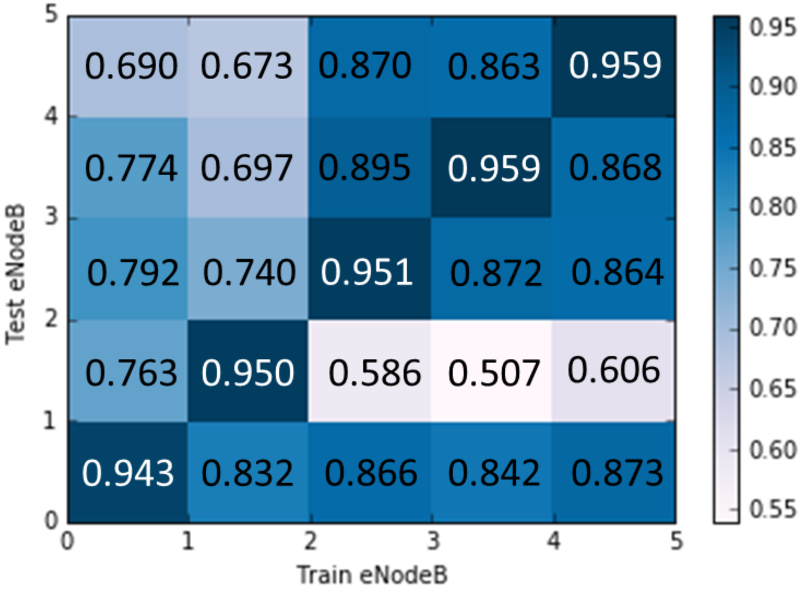
Sample size
|R| = 10…20
for DTW



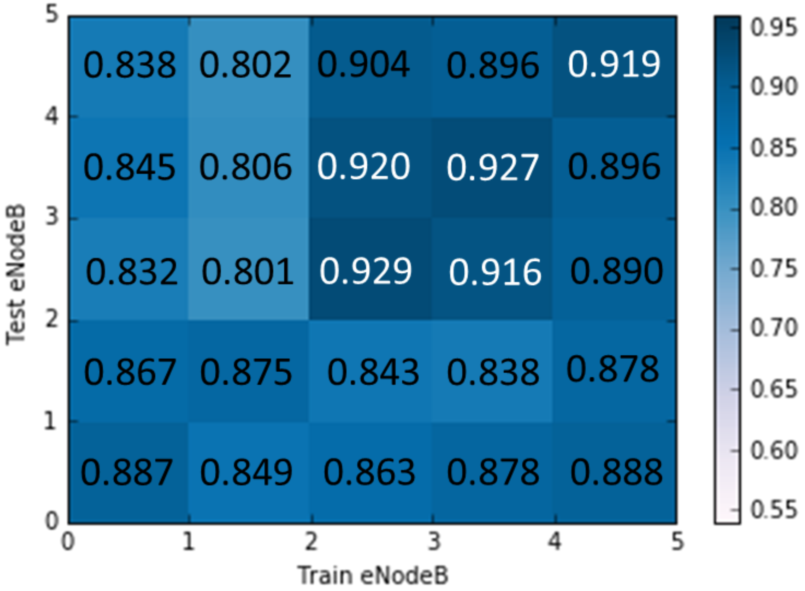
Train and test on different base stations
5 stations with different traffic volume

ERICSSON

MTA SZTAKI



Boosted Tree AUC



Similarity Kernel AUC

Conclusions of the Ericsson use case

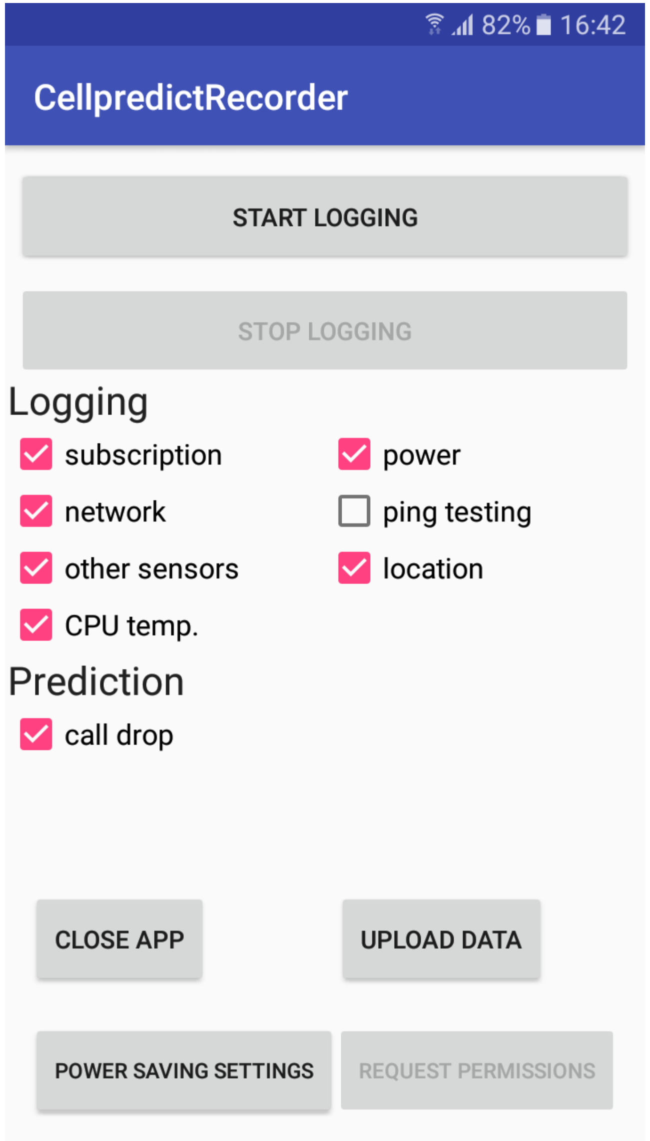
- Reconstruction and justification of known ML techniques (AdaBoost) for new mobile radio technology settings
  - Understanding technological reasons (uplink problems)
  - Assessing the potential of predictions: many seconds ahead, using only short session history
- New, improved methods for multi-time series classification
  - Practically better results
  - New theory based on Fisher kernels applicable for other problems

"Machine Learning Based Session Drop Prediction in LTE Networks and its SON Aspects", IWSON 2015

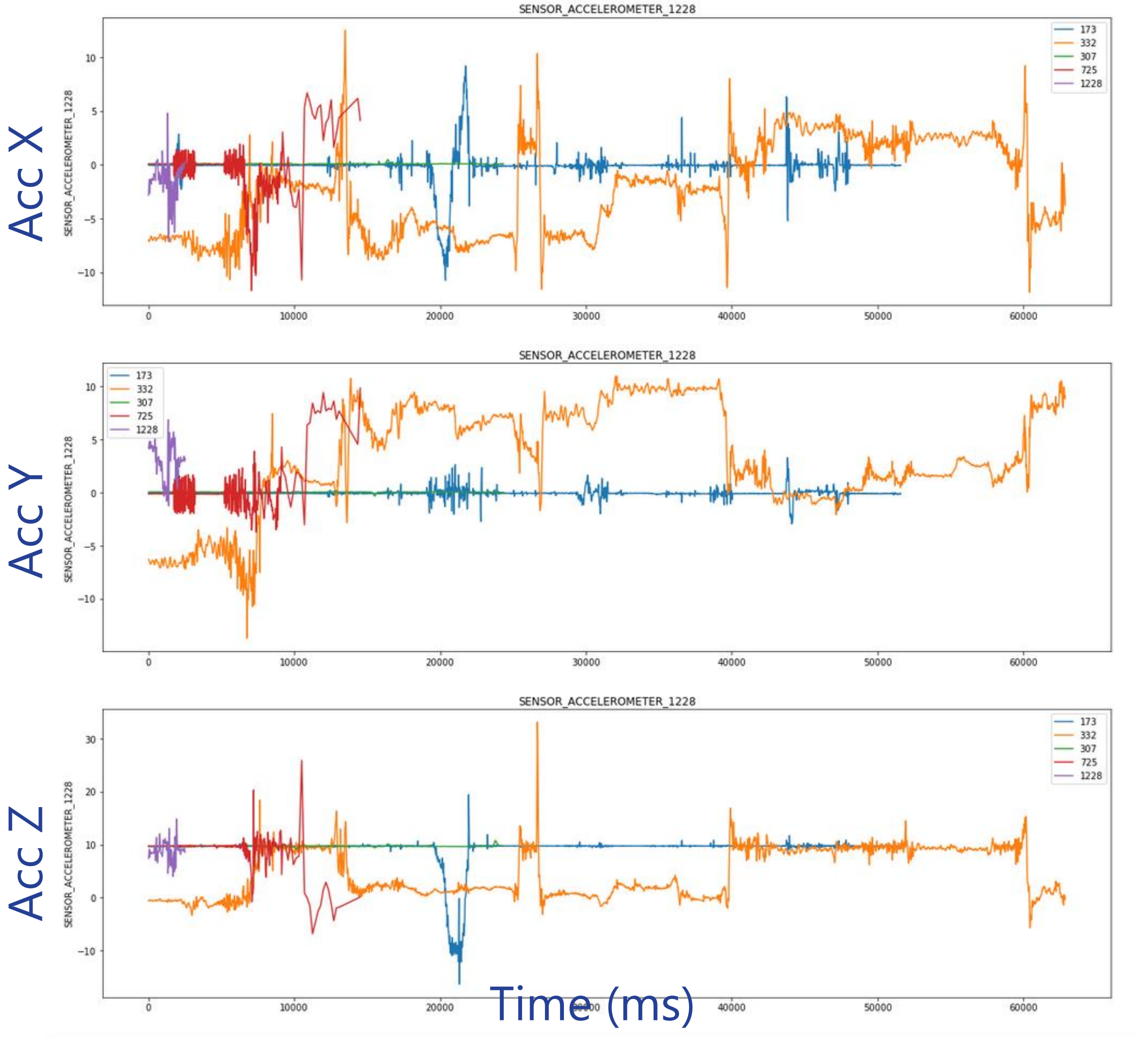
Use case 1b:
Session (call) drop by Smartphone logs

- Csaba Sidló
- Mátyás Susits
- Barnabás Balázs

Logging app and time series sample for illustration

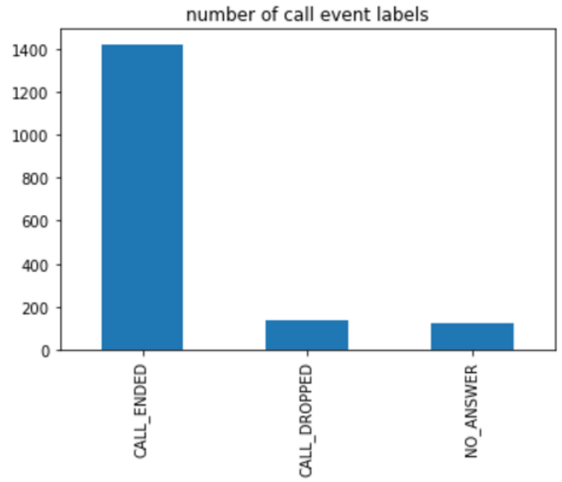


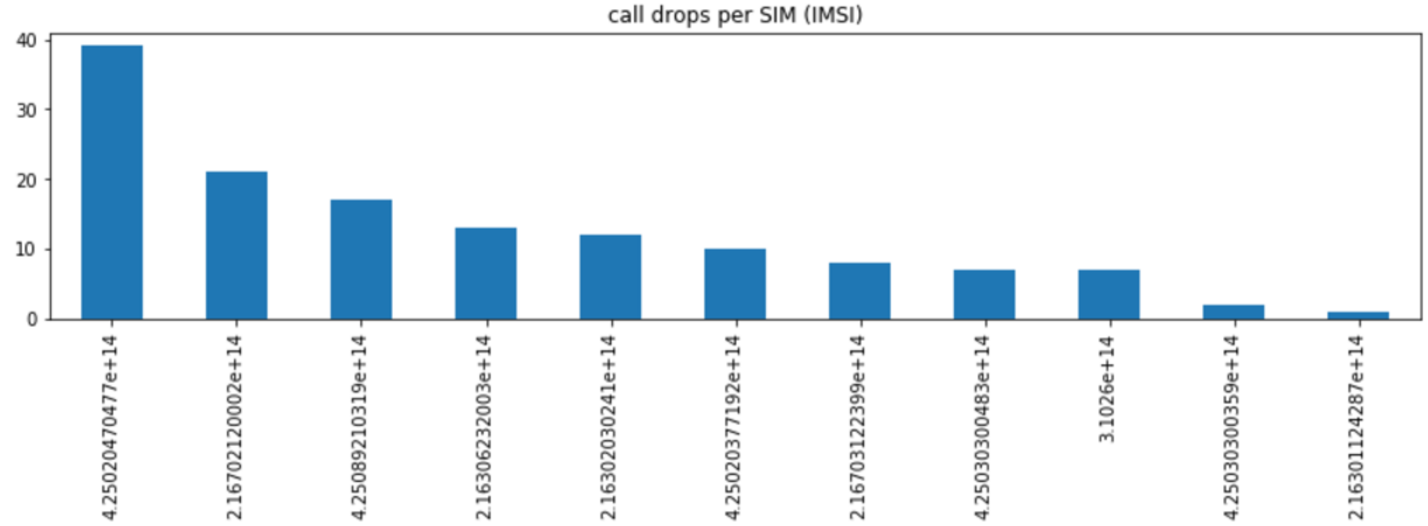
Accelerometer, 5 different devices



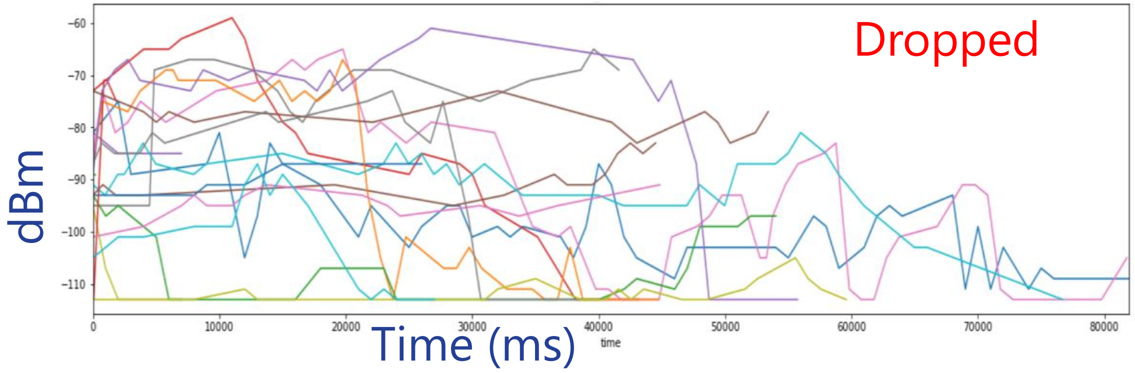
The data: 9 normal and 2 rooted phones (S7 and Motorola G5)

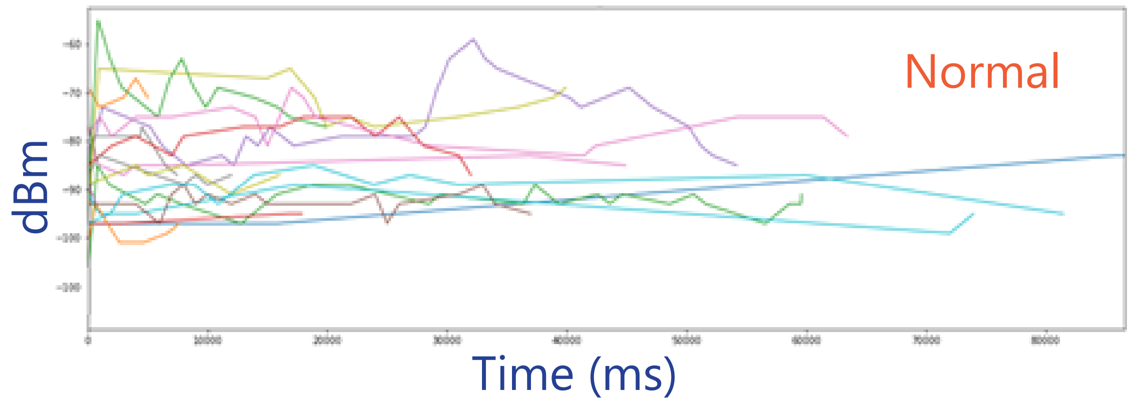
```
CALL_ENDED      1425
CALL_DROPPED     137
NO_ANSWER        125
Name: CALL_LABEL, dtype: int64
```





10-10 sample call signal strength series.
Each line corresponds to a call dropped/terminated at end of line.

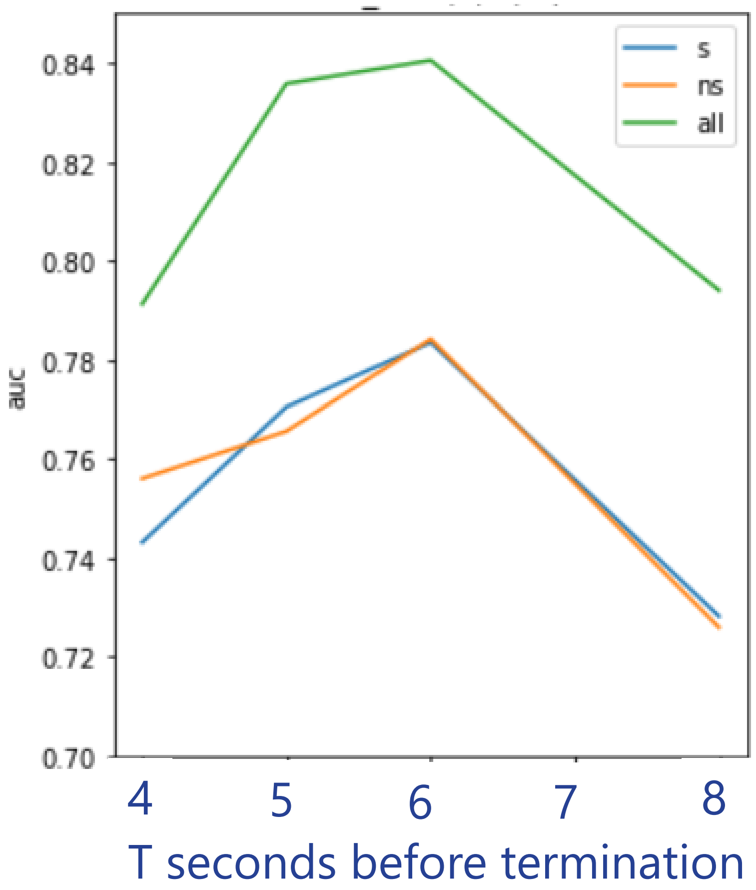




Call drop prediction quality, depending on advance time

MTA SZTAKI

- Measured by AUC
- Using signal strength features (s), all sensors but no signal strength (ns) and all
- Sample setting for timing
  - Last 3 secs of call data not considered
  - Prediction from logs T=4…8 seconds before drop
  - Graph shows AUC as function of T
  - Features in 10 sec sliding window
- Alarm by beep and vibration
  - Features involve drop in quality, which help in reducing unnecessary alarm

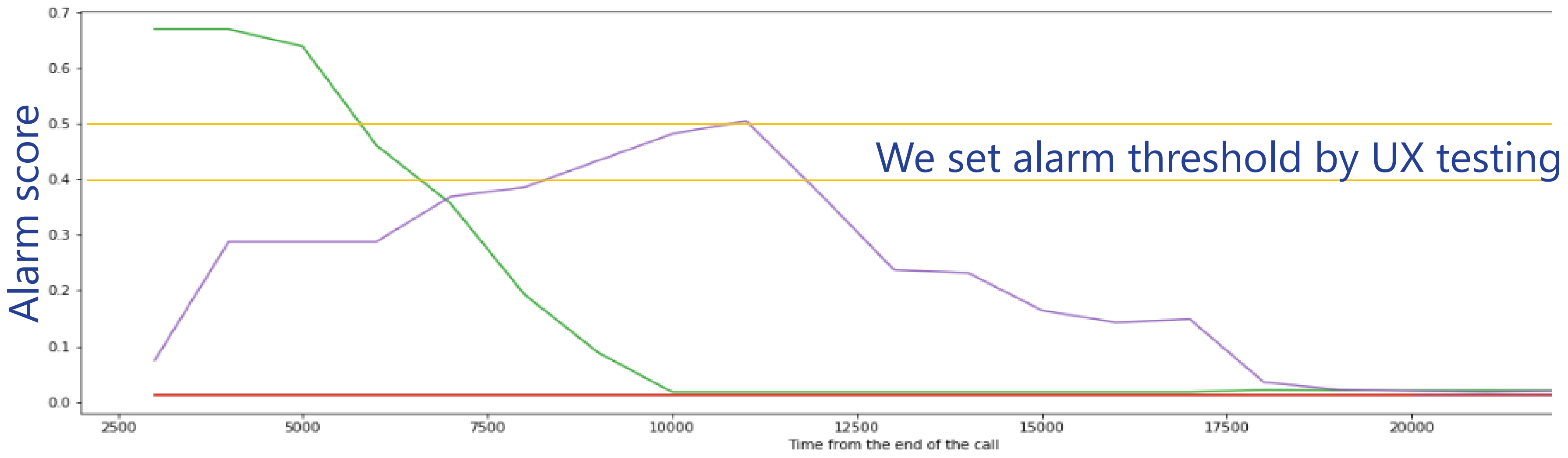


Call drop prediction sample scenarios for dropped calls

MTA SZTAKI

Time measured BACKWARDS from drop

- Green: sudden loss, maybe entering elevator or tunnel
- Purple: low quality throughout the call, not getting worse anymore before drop
- Red: uncaught, probably drop caused on other side (drop real reason only for rooted)



T seconds before termination (time goes backwards)

The model: Boosted Trees

MTA SZTAKI

- Radio signal (dBm) in sliding window
  - Max, min, mean, deviation, max drop
  - Time spent at very low value
  - Several derived features tested and selected by Boosted Trees

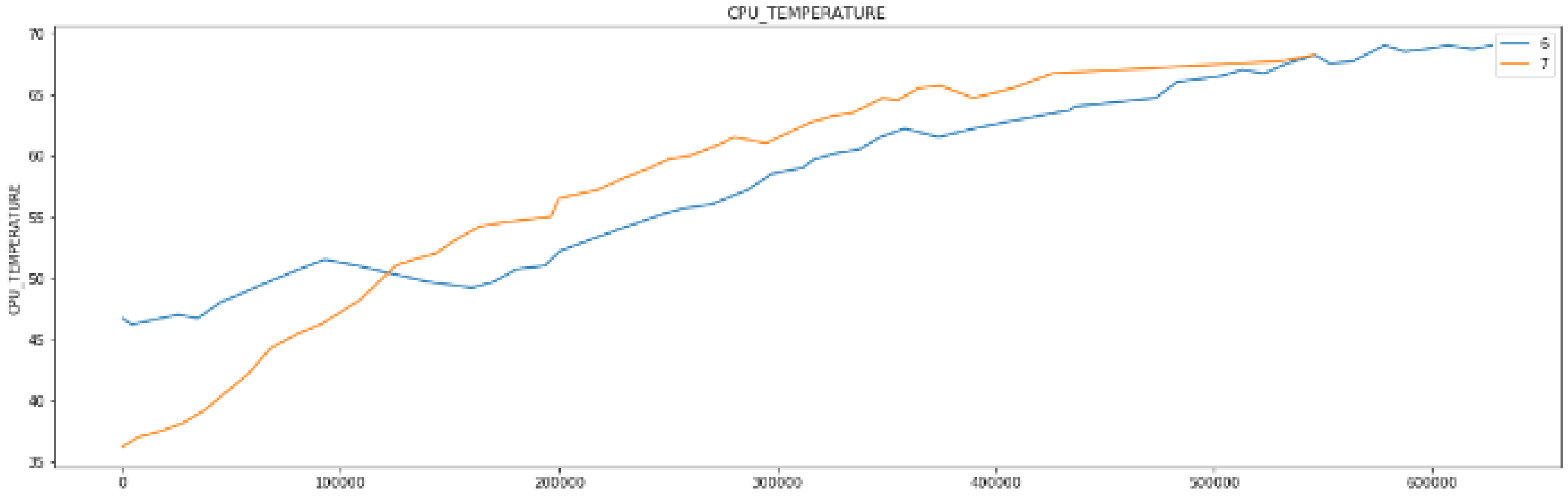
- Sensors
  - Battery charge, magnetic, light, accelerometer
  - Min, max, diff, mean, deviation

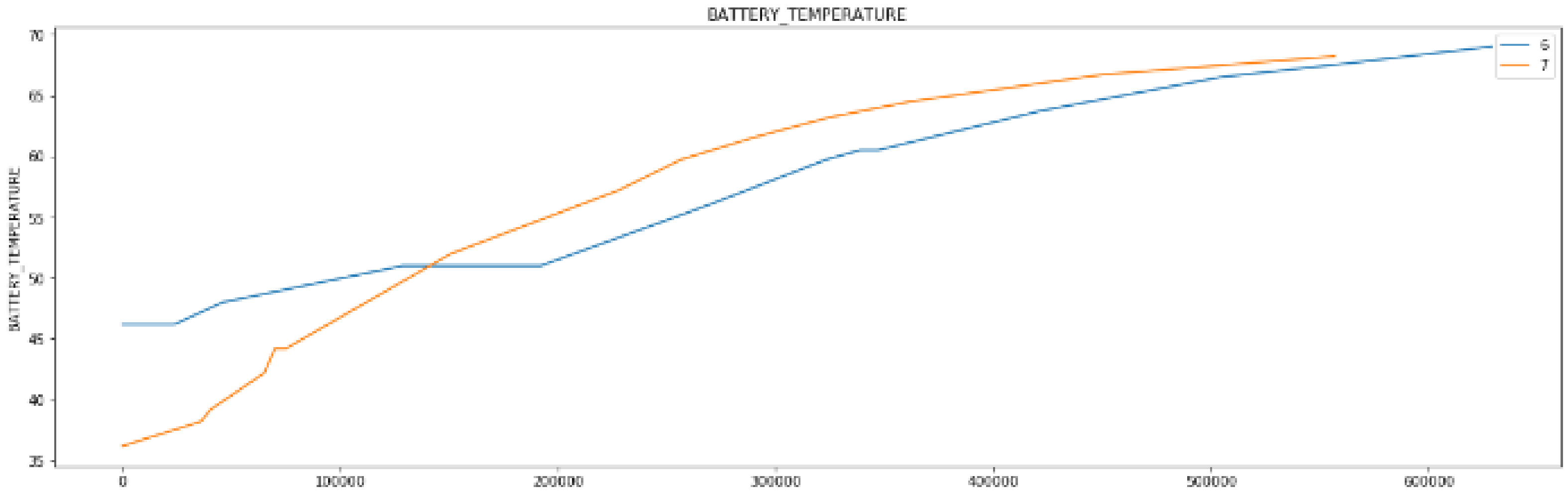
- Most important non-signal-features
  - Magnetic changes
  - Motion
  - Getting darker
  - Low battery (power save mode?)

Other reasons? Overheated phone test until turn off (experimentation to be continued)

MTA SZTAKI

Heating until drop
(by CPU load + lamp)
Drop at 70C

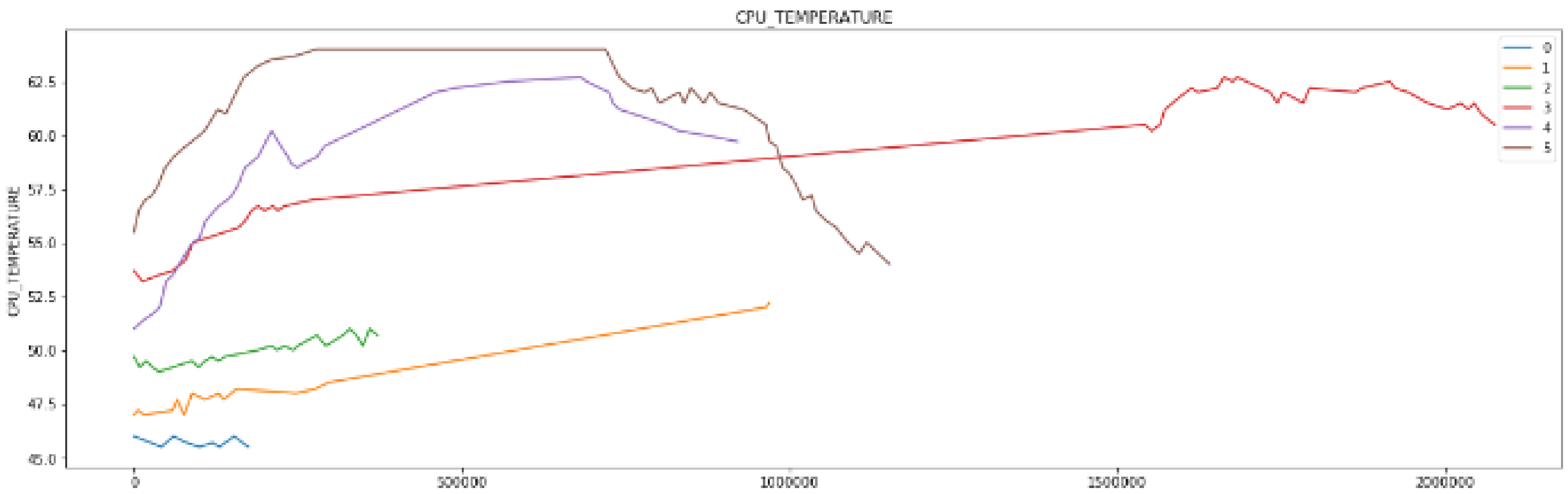


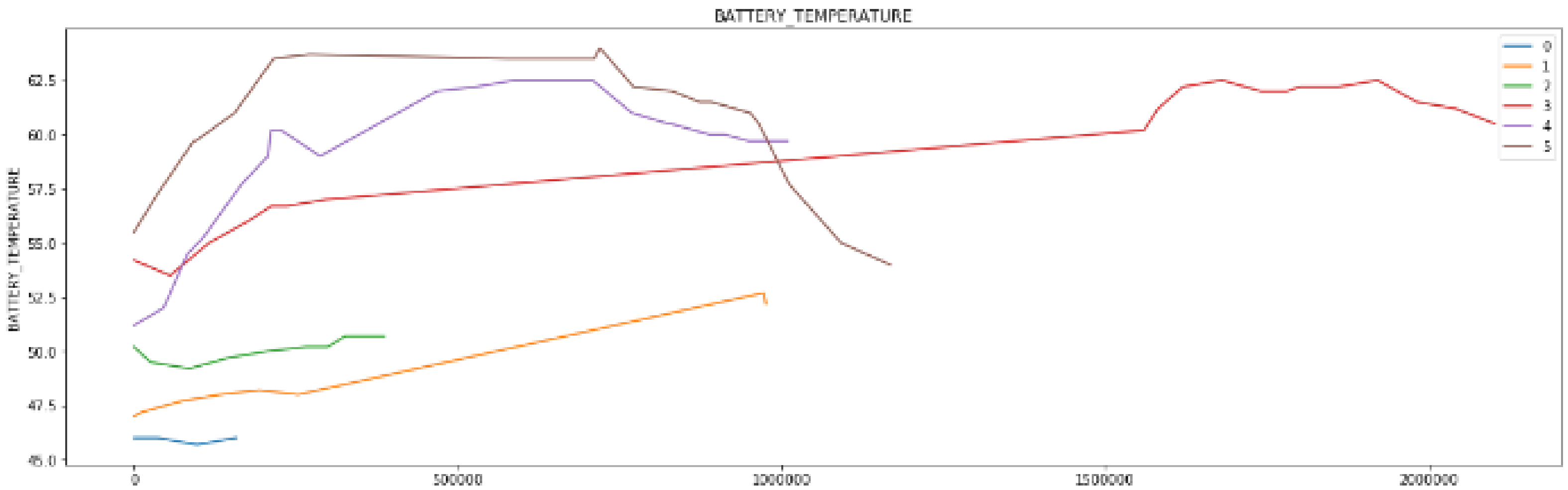


Other reasons? Overheated phone test until turn off (experimentation to be continued)

MTA SZTAKI

Normal calls still
up to 65C





감사합니다!

Thank You!

תודה!

MTA SZTAKI
Hungarian Academy of Sciences
Institute for Computer Science and Control

Use case: scrap rate prediction in transfer molding

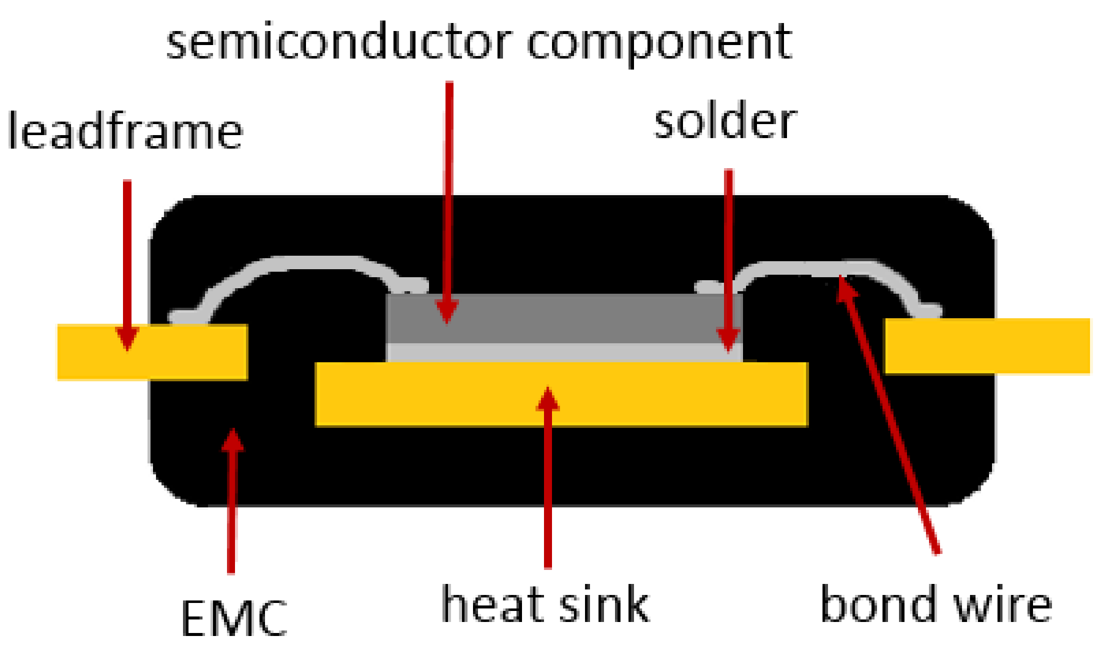
BOSCH

Molding electronic components

- Electronic components are packaged in plastic to prevent from external damage
- Part of the manufacturing process:

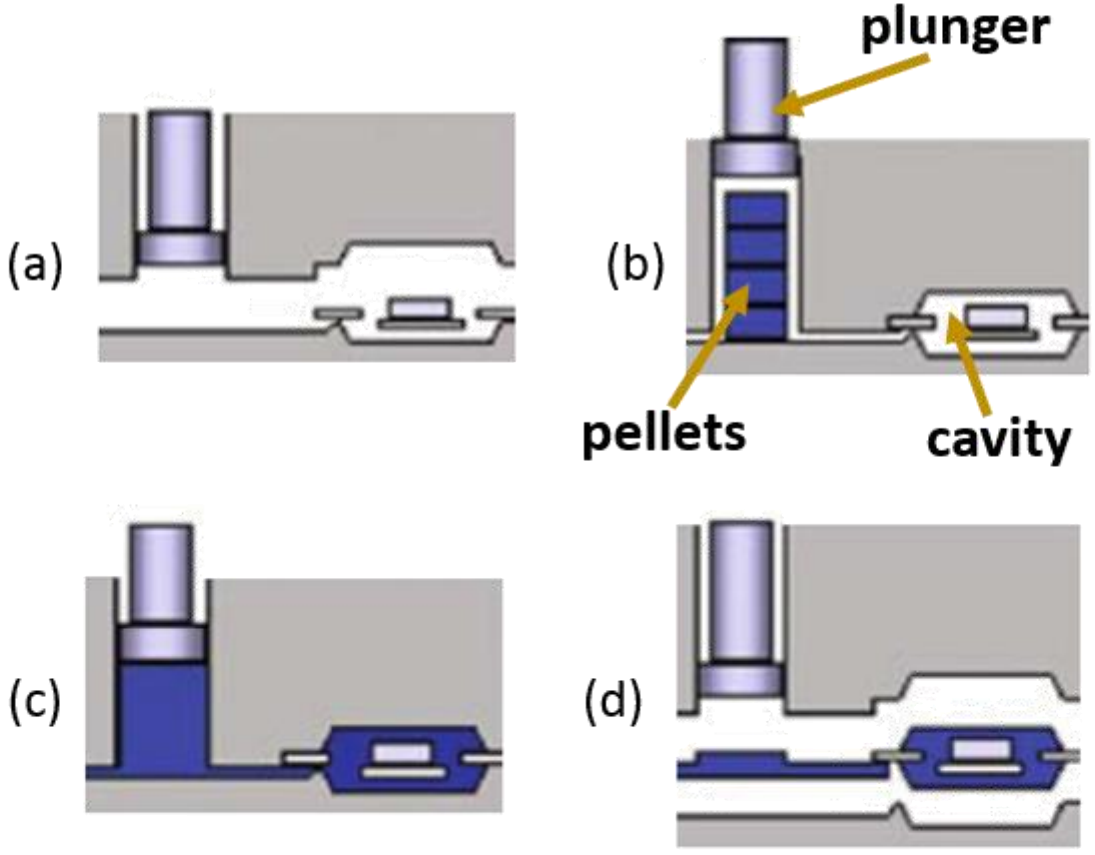
  …
  1. Soldering to heat sink
  2. Degreasing
  3. Bonding on lead frame
  4. **Molding**
  5. Inspection (ultrasonar or visual)

  …
- Possible failures (scrap)
  - Void: air inside
  - Delamination
  - Cracks

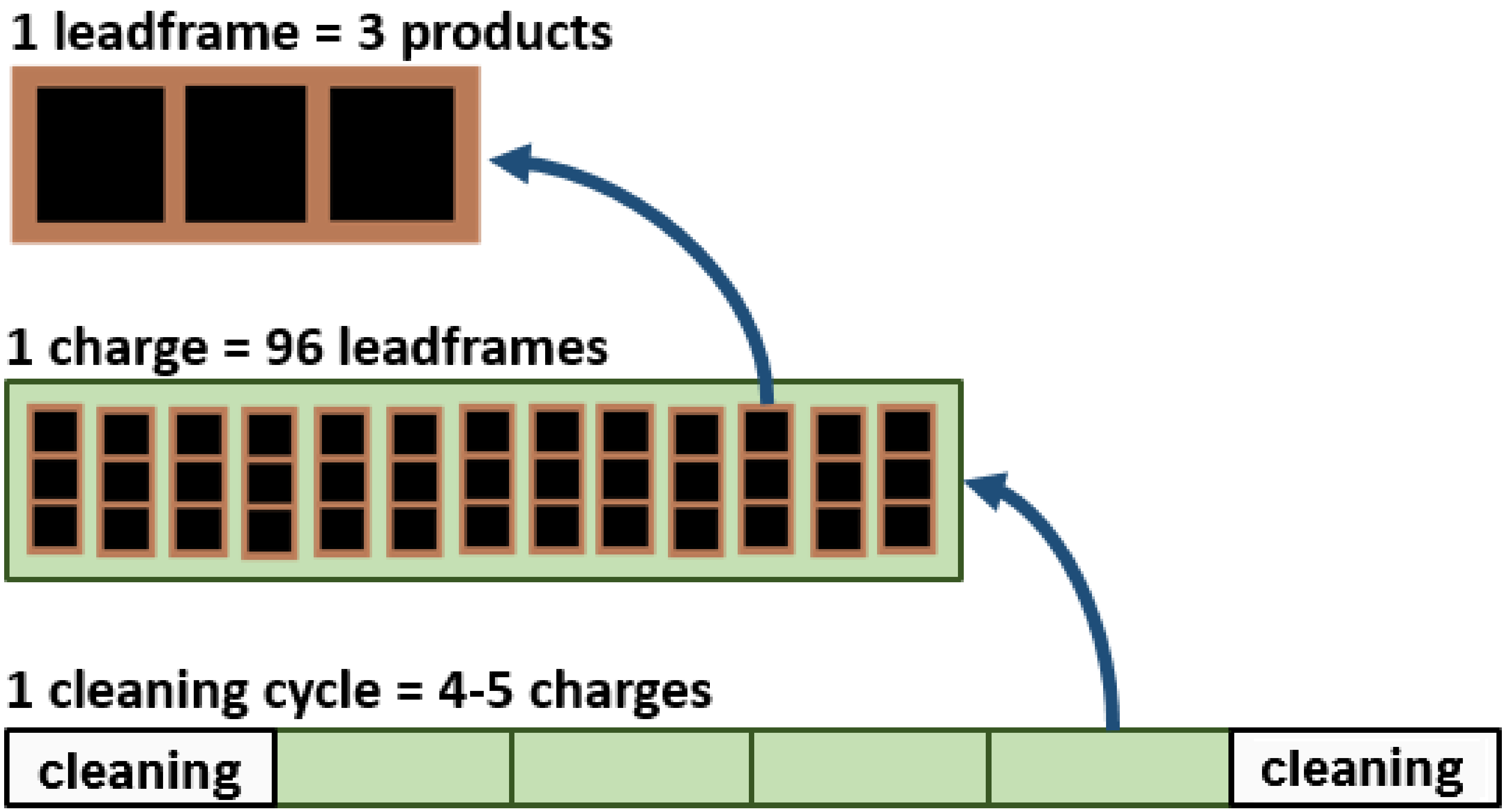


Transfer molding

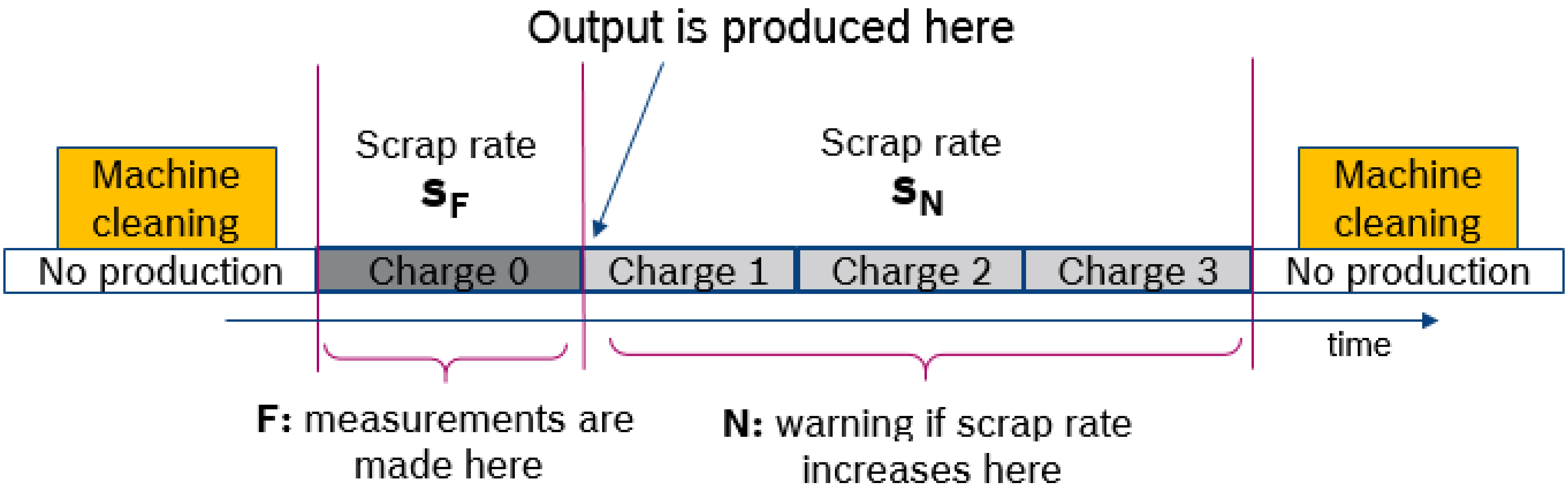
- Automold machine
- Four presses
- Two cavities in each press
- Leadframes have a few products (3 in our case)
- **Shot**
  - Smallest unit of production
  - Two leadframes are produced, one in each cavity
- **Charge**
  - No production parameters are changed
  - 96 leadframes, 288 products
- Daily shift
  - Roughly 10 charges
  - Different products may also be produced in different charges



Cleaning cycles

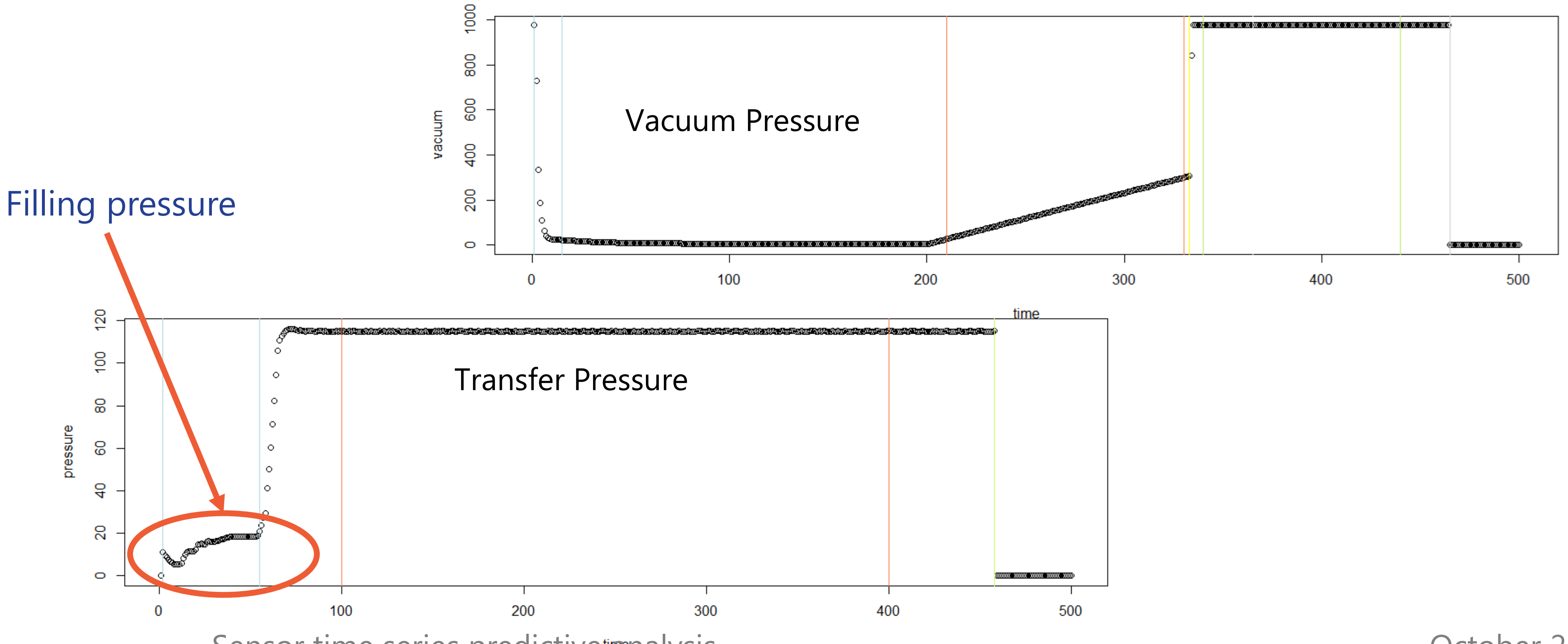


Scrap rate prediction task

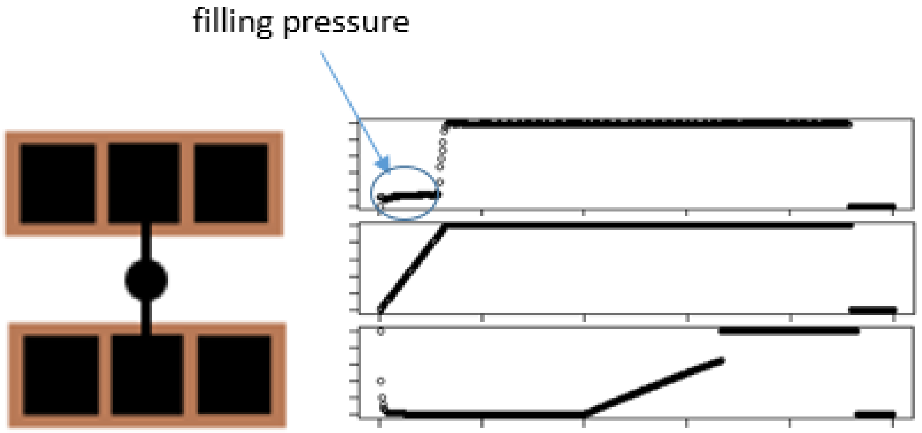


Data

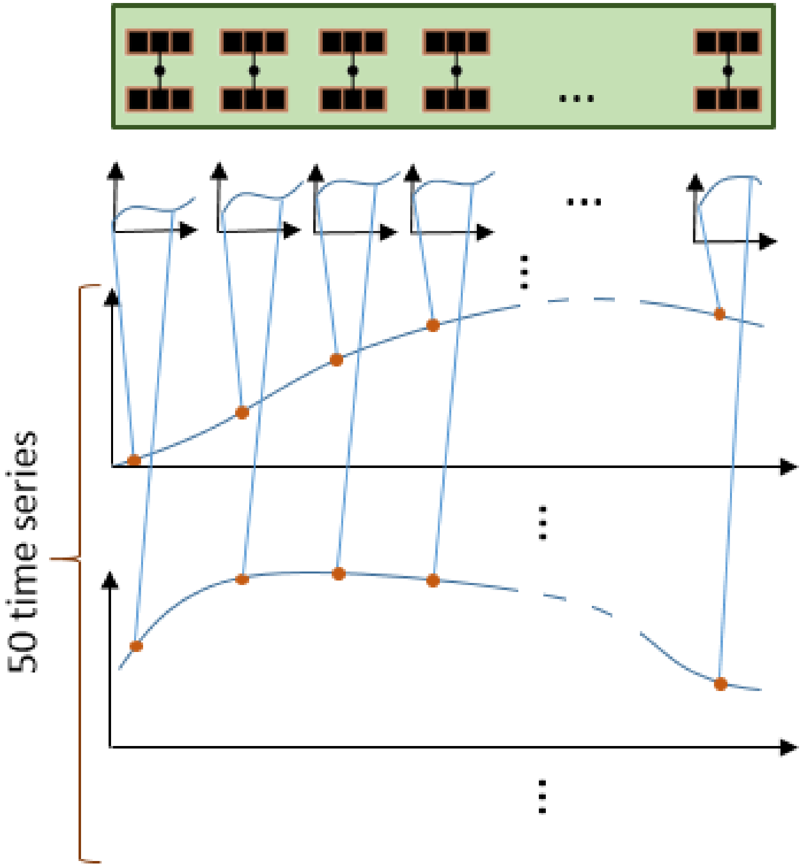
- Multiple time series, 500+ measurements in each: many, many data points
- Features of mean, variance, differentials, distribution of transfer graph data points
- In the end, ~50 features in the final classification data



Time series of time series



each shot 3 transfergraphs are recorded



many shots are made in each charge

filling pressures for each shot

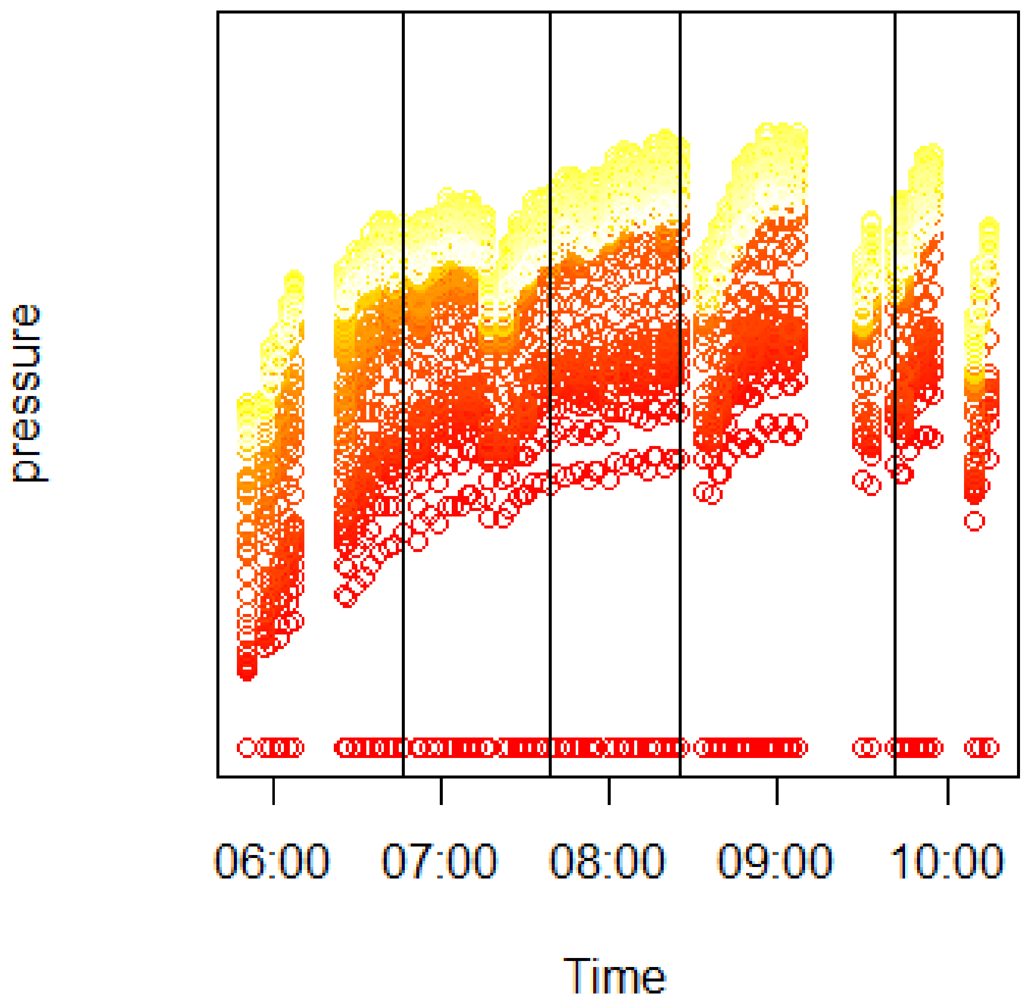
time series of the 1st measurements

⋮

time series of the i. measurements

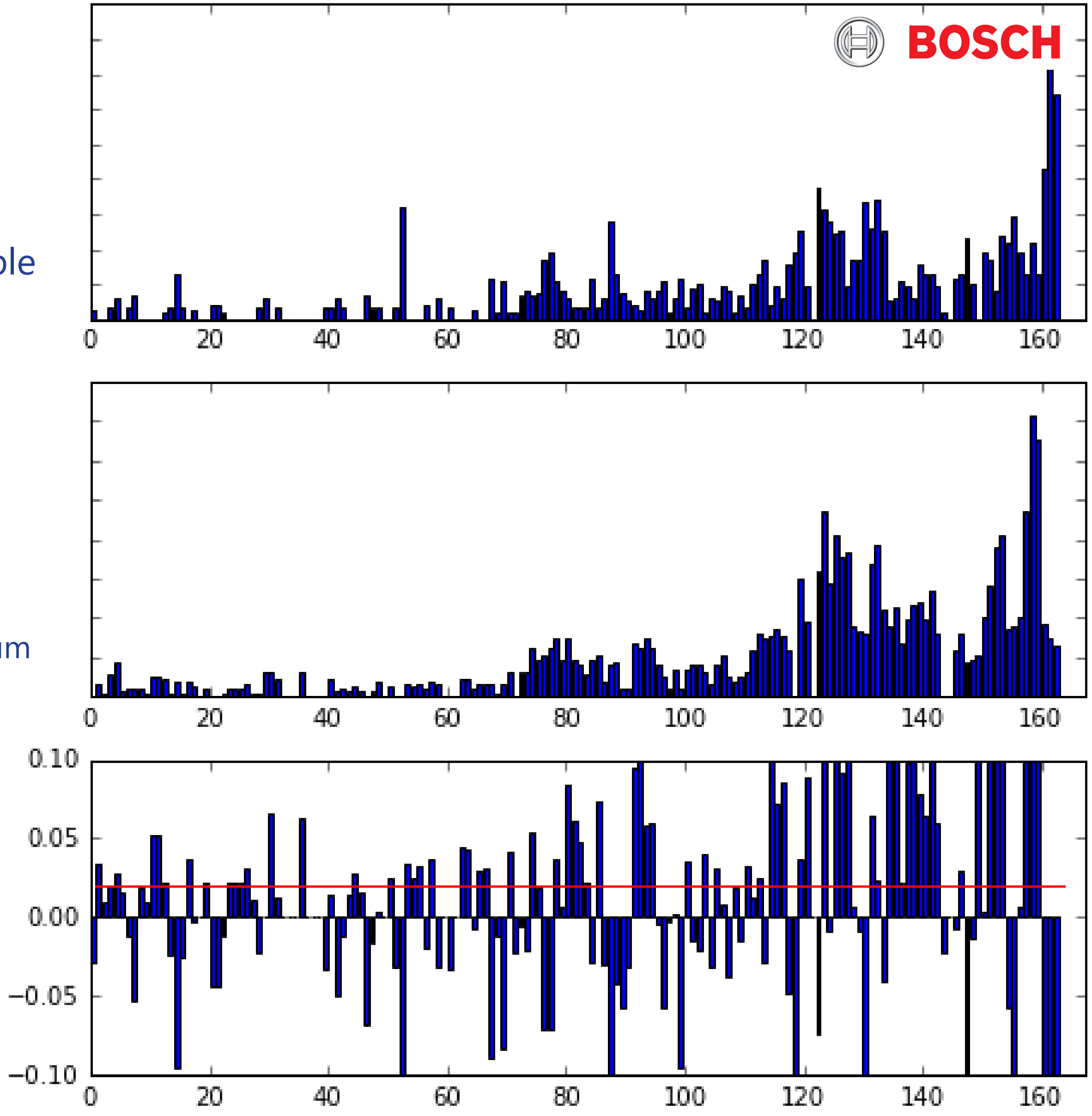
Series of pressure series

- Filling pressure measurements
- Red to yellow color scale: measurement 1 – 50
- Vertical lines: new charges
- **Important to notice**
  - Values are lower after machine cleaning, and later they typically increase
  - Pressure is indirectly measured on the plunger
  - Contamination modifies pressure measurement



Prediction target

- Three different periods
  - First with low scrap rates
  - Second with intermediate but still acceptable scrap rates
  - Last with very high, inacceptable
- In the same period, we will quickly learn the average scrap rate
  - Removed all features that may refer to the production period
    - External air pressure (vacuum max)
    - Vacuum transfer slope increases slowly as vacuum chamber quality decreases, yearly maintenance only
    - Pellet size only changed once in our period
- TARGET: predict increase over previous period
  - Multiplicative 5% AND
  - Additive 2%
  - Works for both low and high rates



MTA SZTAKI

F: scrap rate in first charge after cleaning

N: average scrap rate in next charges

$$1.05 \cdot s_F - s_N$$

Prediction for individual products?

MTA SZTAKI

- Makes no sense, since the product is immediately inspected after released
  - We know scrap as soon as we have the data
  - Prediction BEFORE the data, based on past trends
- Makes sense for ROOT CAUSE ANALYSIS
  - Why did the product fail
  - How should we change parameters
  - Feature importance, thresholds are more important than model prediction

Feature importance

- Calculated by Gradient Boosted Trees
  - In each iteration, a small decision tree is computed
  - Next trees refine on the error, decision is independent of the previous tree
  - Importance is based on how early and how frequently a variable is used in the trees
- Important for individual product scrap prediction and used for the rest of the analysis
  - Filling pressure curve
  - Vacuum curve
  - Shot count (time since cleaning)
  - Time elapsed after degreasing

Modeling methods

- Baselines: just consider past scrap rates
  - No new information for the operator
  - Yet, we should be at least this good
- Feature based methods
  - Minimum, maximum, mean, median, variance
  - Gradient Boosted Trees (GBT) – considered best for general mixed data type problems

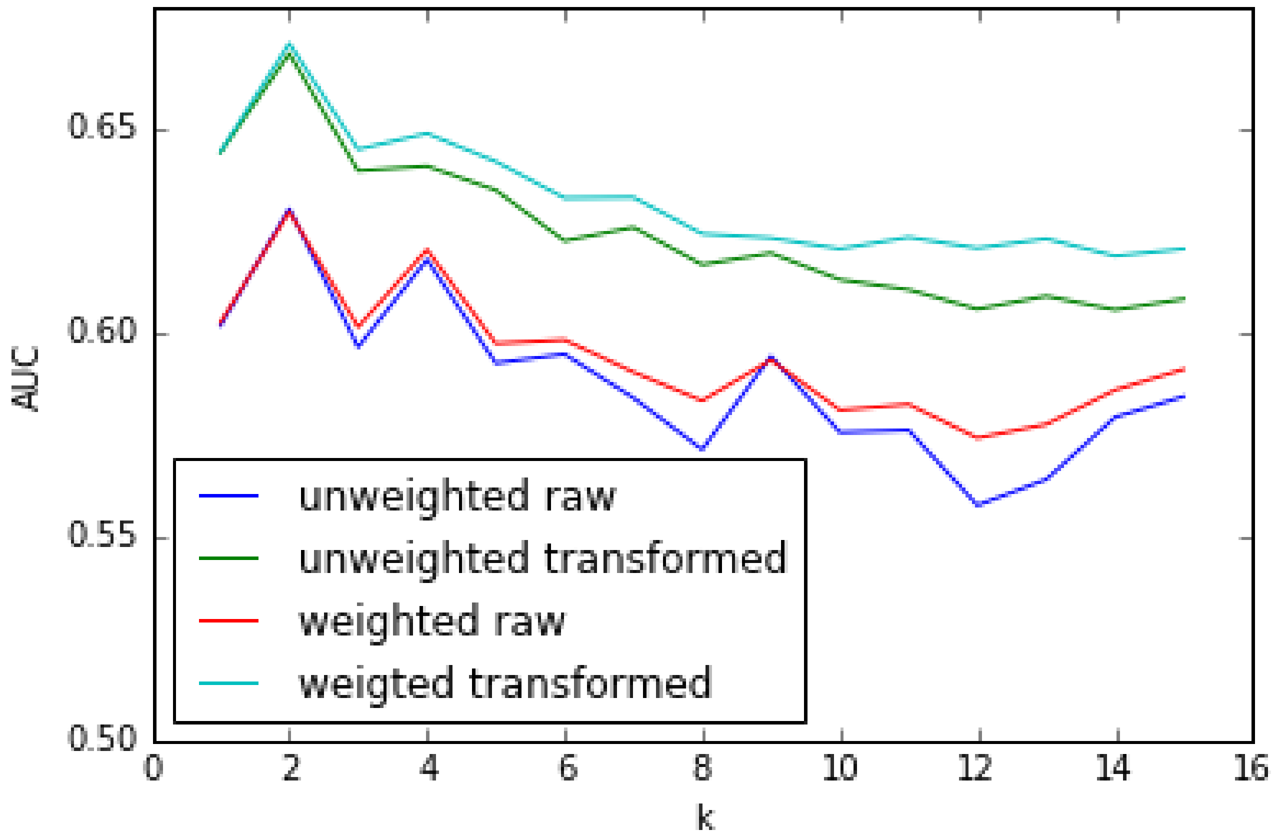
Shape based methods

- Nearest neighbor k-NN
  - 50 distance values
  - Use Euclidean length to determine most similar charges
  - Parameter to consider the k nearest data points
- Support Vector Machines
  - Set aside a sample of past charges to define variables as distance from these charges
  - Classification: SVC
  - Regression: SVR

Nearest neighbors (k-NN) as function of k

- Unweighted:
  - Different models after each first charge
  - All models produce a score, either binary likelihood or scrap rate regressor
  - Recall target: $1.05 \cdot s_F - s_N > 0.02$
- Unweighted
  - Raw score directly used across testing charges
- Weighted
  - Transform raw scores first by

$$\text{pred} - 1.05 \cdot s_F$$



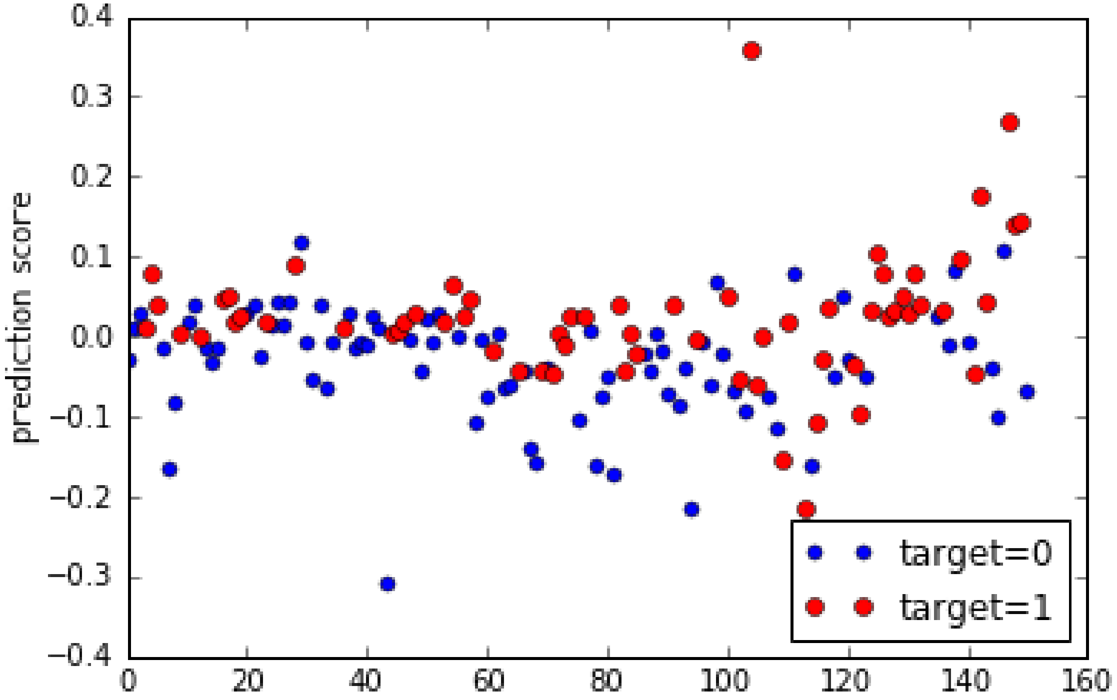
Results

MTA SZTAKI

Method	AUC
SVR on features (weighted)	0.70
SVR on scrap rates (weighted)	0.68
2-NN (weighted)	0.67
2-NN	0.63
GBT on scrap rates	0.61
SVC	0.59
SVR on scrap rates	0.59
SVR on features	0.59
GBT on features	0.54

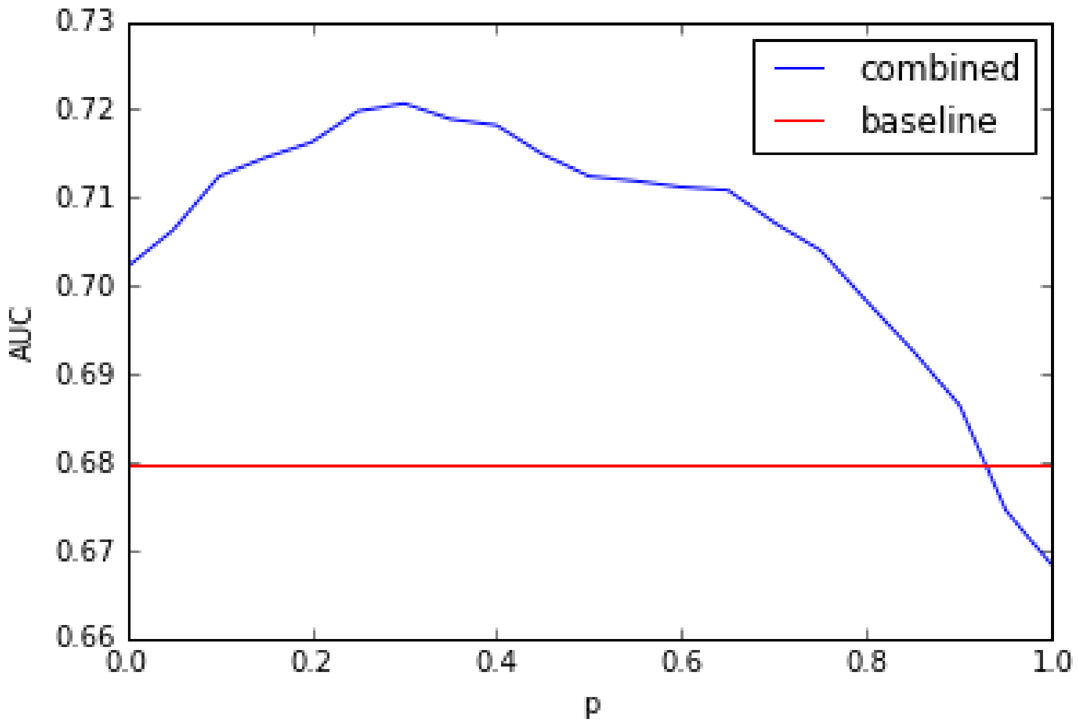
Results

Target events and prediction scores as the function of time (cleaning cycles 1-157)



Linear combination with weight p

- SVR on features (p=0)
- 2-nearest neighbor (p=1)



Summary of the Transfer Molding use case

BOSCH MTA SZTAKI

Facts:

6-10,000 products/hour
Few 100 data points per product
100> failed products in a day

Available for several months:

Delamination, Void, … failures
- First line reject
- Second line reject
- Failures at later stages

Machine logs

- BottomPreheatTemperature1-6
- BottomToolTemperature
- UpperToolTemperature
- LoaderTemperature
- PreheatTime
- toolData[1-2].value
- TransferPressGraphPos1-150
- TransferSpeed1-10
- TransferTime
- TransferVacuum

Root Cause Analysis

Transfer pressure measured indirectly
- includes friction from valves
- certain points in the time series indicate contamination

Vacuum pressure has no direct effect, but …
- Differences in trays: calibration problems
- Vacuum drop speed: leakage and blinding
- May result in less effective cleaning?

Result in variables that affect the production in indirect way that needs to be understood

Summary

- Physical processes produce rich structured data
- Mix of time series and static data is challenging to classify
- Time series of time series is even more
- Combination of feature and shape based methods
- Fisher kernel is a promising new direction

Questions?

András Benczúr
Institute for Computer Science and Control (MTA SZTAKI)
Hungarian Academy of Sciences
benczur@sztaki.mta.hu http://datamining.sztaki.hu/