Human-in-the-Loop Large-Scale Predictive Maintenance of Workstations


Alexander Nikitin Samuel Kaski

FCAI Finnish Center for Artificial Intelligence
Aalto University
MANCHESTER 1824 The University of Manchester

TL;DR

- We implement a human-in-the-loop (HITL) predictive maintenance (PdM) system for workstations
- Predictive maintenance: Predict problems beforehand
- Human-in-the-loop: Maintenance engineers give their decision rules, which comes naturally to them
- We show effectiveness on synthetic data, real historical data, and in field-test

Introduction

Workstation PdM: Fix issues in a large number of workstations without user interruptions.

Why is this difficult?

- Thousands of workstations and dozens of customers
- Highly varying load of experts over time
- Assist, not center workflow
- The service needs to be reliable

What is missing?

- Include both predictive algorithms and expert knowledge
- Field-tested PdM algorithms

Solution. We implement a machine learning system that predicts future problems in sets of workstations (computers, laptops, and servers). Domain experts are included in the loop not only as providers of correct labels, as in traditional active learning but as a source of explicit decision rule feedback.

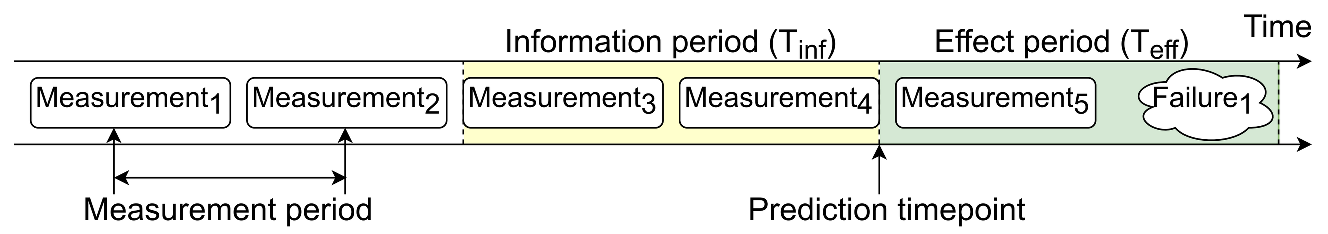
Methods

Algorithm overview

(A) Beforehand, on historical data:
  (i) Optimize data preparation hyperparameters
  (ii) Learn a model to predict the probability of a problem

(B) Iteratively, during operation:
  (i) Elicit expert feedback as decision rules
  (ii) Use ML model and elicited decision rules to predict future problems

Data Preparation



- average features using aggregators over T_{inf}
- formalize failure: number of alerts is larger than a threshold
- predict the probability of failure in T_{eff}

To extract features, we use aggregation operations from $\mathcal{A}$:

$$\tilde{\boldsymbol{x}}_i = \bigoplus_{\text{agg}\in\mathcal{A}} \text{agg}\left(\boldsymbol{x}_i^{t-T_{\text{inf}}}, \ldots, \boldsymbol{x}_i^{t}\right). \tag{1}$$

Bayesian Optimization (BO) of Hyperparameters

We employ BO to optimize data preparation hyperparameters (incl., T_{inf} and T_{eff}). We use GP as a surrogate for the model's predictive performance. A Gaussian process (GP) is a prior over functions:

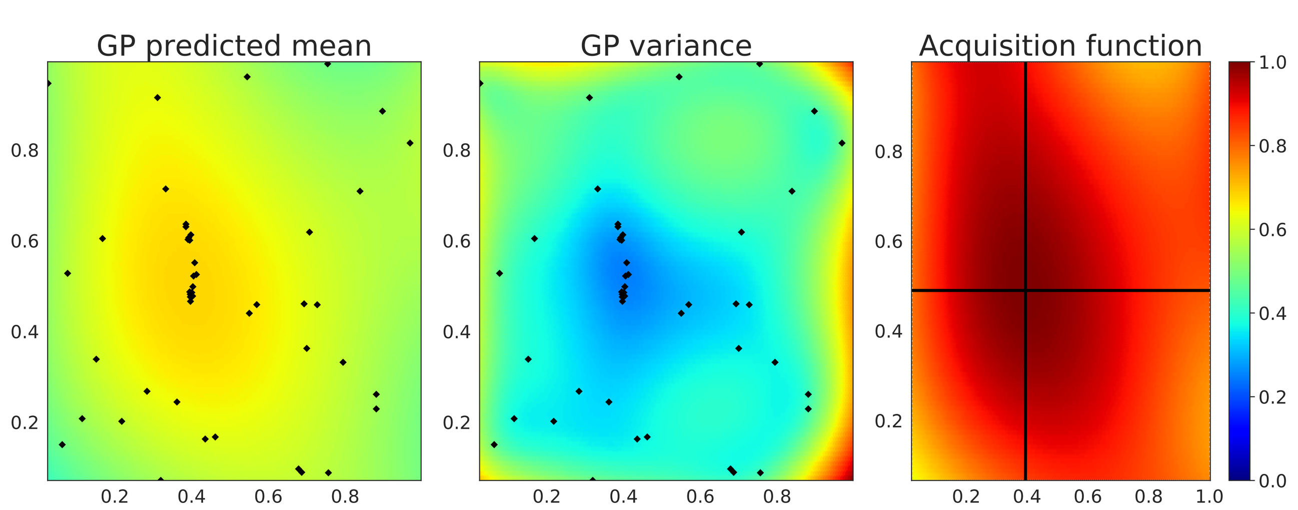
$$f \sim \mathcal{GP}(\mu(\mathbf{x}), K(\mathbf{x}, \mathbf{x})). \tag{2}$$

We used the RBF kernel:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2), \tag{3}$$

where γ is a length scale hyperparameter. Acquisition function:

$$a_{UCB} = m(\mathbf{x}) + \kappa\sigma(\mathbf{x}), \tag{4}$$



Humain-in-the-loop approach

(i) Experts can use their knowledge to improve predictions, and (ii) the model's decisions are explainable.

Decision rule elicitation

- Data-driven machine learning model M,
- Experts can explain their decision-making process via heuristics,
- the heuristics can be represented as simple decision rules,

Model based on decision rule feedback:

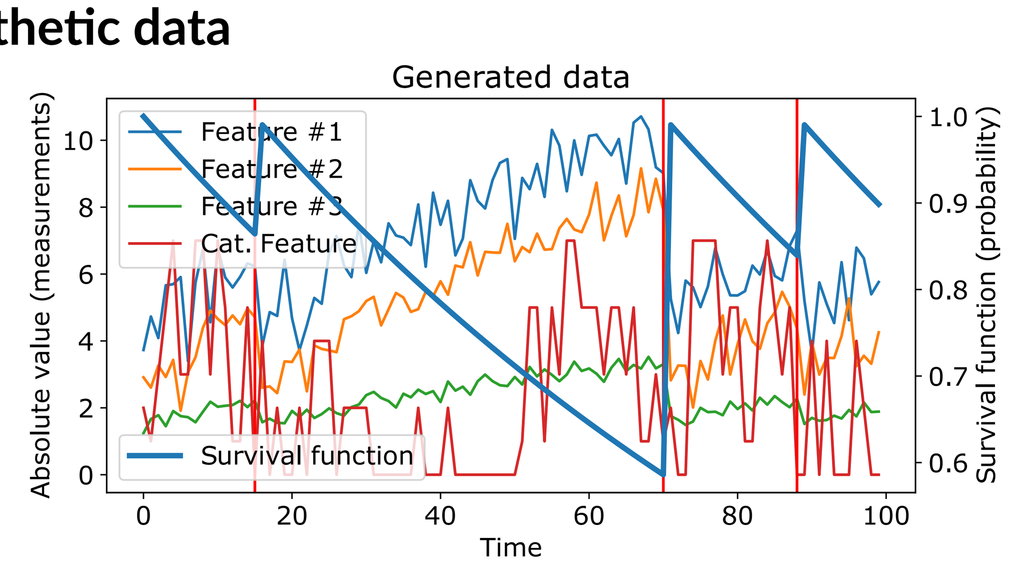
$$C_{\text{fb}}(\boldsymbol{x}) = \zeta\Big(\sum_{i=1}^{F} \overbrace{f_i(\boldsymbol{x})}^{\text{Elicited decision rules}} \underbrace{\text{sim}(\boldsymbol{X}^i_{\text{test}}, \boldsymbol{x})}_{\text{Similarity}} \overbrace{\theta_i}^{\text{Weight}}\Big). \tag{5}$$

The whole model:

$$C(\boldsymbol{x}) = \overbrace{\alpha}^{\alpha\in[0,1]} \underbrace{\text{M}(\boldsymbol{x}, \boldsymbol{\theta}_M)}_{\text{ML model parameterized by } \theta_M} + (1-\alpha)\overbrace{C_{\text{fb}}(\boldsymbol{x}, \boldsymbol{\theta}_f)}^{\text{Feedback model}}, \tag{6}$$

Experiments

1. Synthetic data

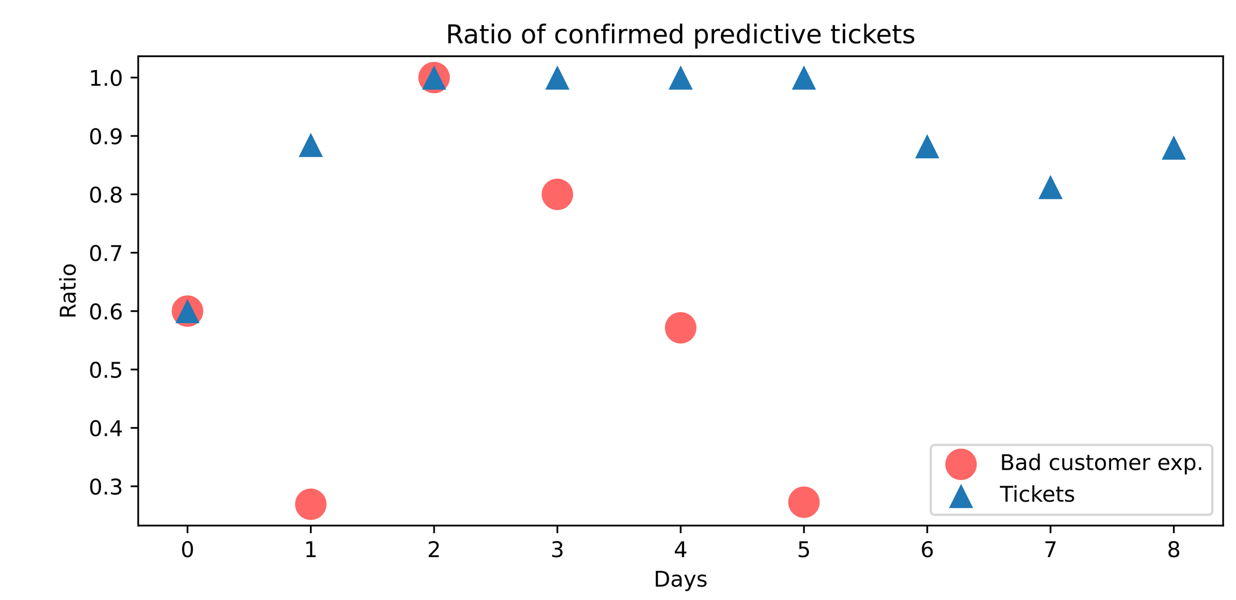


2. Real historical data

Algorithm	f_1-score	precision	recall
Logistic regression	0.51 ± 0.02	0.88 ± 0.02	0.35 ± 0.01
ExtraTrees classifier	0.76 ± 0.01	0.72 ± 0.02	$\mathbf{0.82 \pm 0.01}$
Gradient boosting	0.77 ± 0.01	0.87 ± 0.02	0.69 ± 0.02
DRE (3)	0.77 ± 0.01	$\mathbf{0.89 \pm 0.02}$	0.67 ± 0.02
DRE (5)	0.78 ± 0.01	0.87 ± 0.02	0.71 ± 0.02
DRE (15)	0.80 ± 0.01	0.87 ± 0.02	0.74 ± 0.02
DRE (20)	$\mathbf{0.81 \pm 0.01}$	0.84 ± 0.02	0.77 ± 0.02

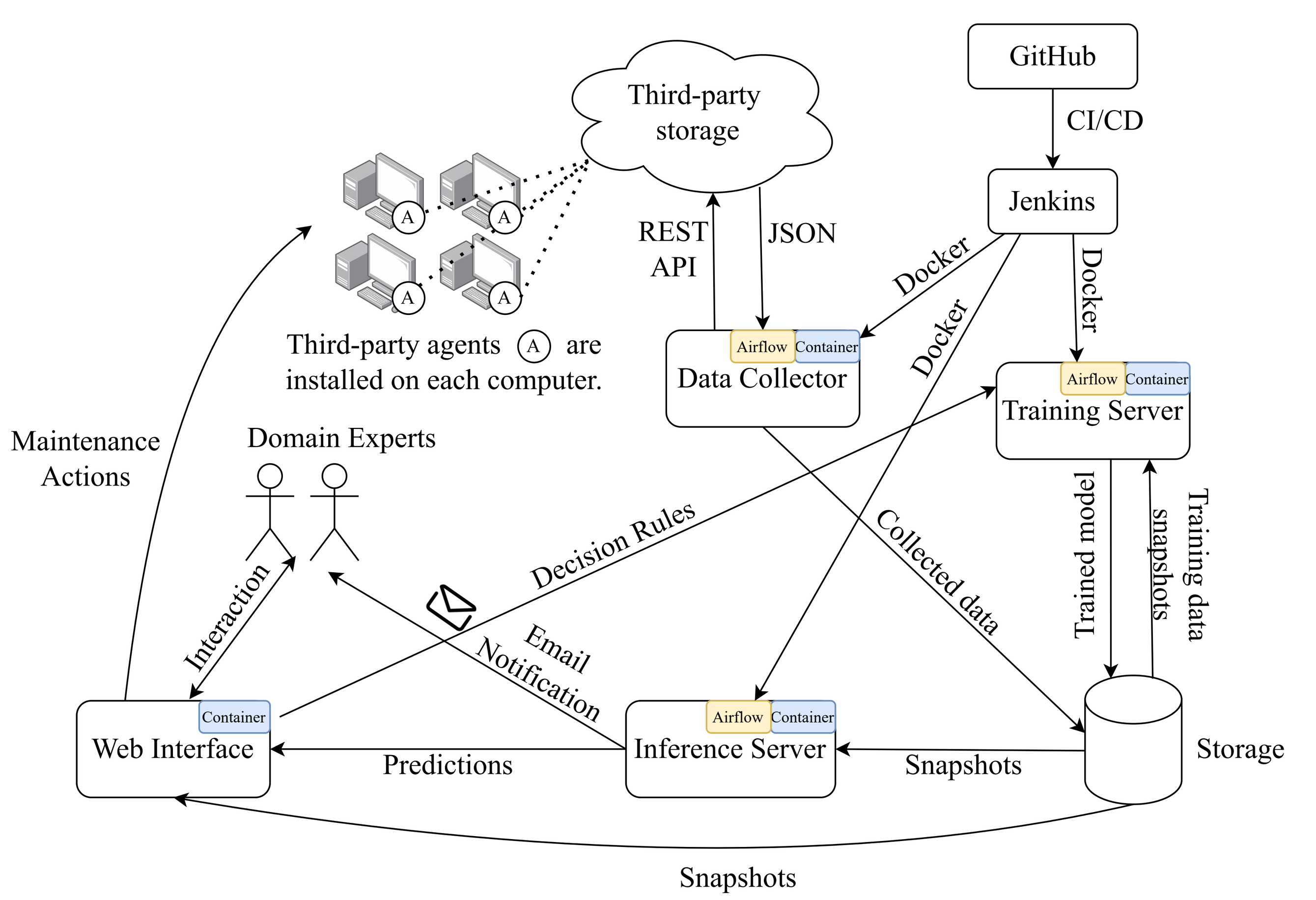
$T_{\text{inf}} = 31.9\text{h}$ and $T_{\text{eff}} = 119.9\text{h}$

3. Field-testing



Deployment

- Services: data collector, training server, inference server, and web interface
- domain experts use email notifications and web interface
- the collected data are stored in an S3 compatible storage
- Used tools: Kubernetes, Airflow, Jenkins, and Ceph



Bibliography

Andreas Holzinger. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*, 3(2):119–131, 2016.

Alexander Nikitin and Samuel Kaski. Decision rule elicitation for domain adaptation. In *26th International Conference on Intelligent User Interfaces*, pages 244–248, New York, NY, USA, 2021. Association for Computing Machinery.

alexander.nikitin@aalto.fi

https://arxiv.org/pdf/2206.11574.pdf