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


Alexander Nikitin[1] Samuel Kaski[1,2]

[1]Aalto University, Finland
[2]University of Manchester, UK

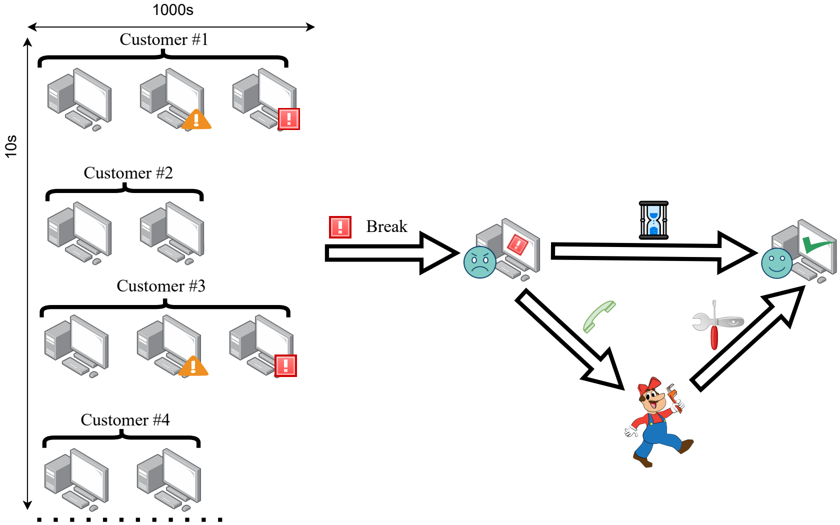
Aalto University School of Science

MANCHESTER 1824 The University of Manchester

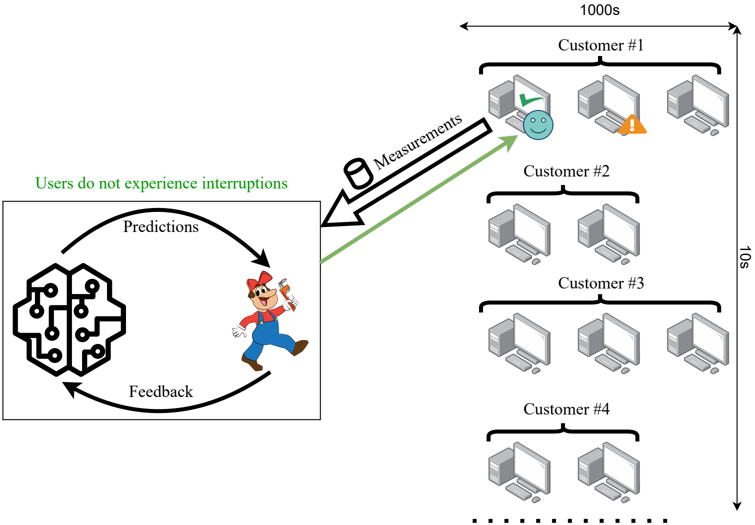
KDD 2022

acm

Workstation maintenance



Proposed solution:
Human-in-the-loop predictive maintenance (PdM)



PdM is coupled with ML challenges

Why is this difficult?

- Thousands of workstations and dozens of customers
- Highly varying load of experts
- Help, not center workflow
- The service reliability

What is missing?

- Include both predictive algorithms and experts
- Field-tested PdM algorithms

Elements of the proposed solution

- Bayesian optimization for parameter search
- Predictive modeling via ensembles
- Decision rule elicitation
- Deployed infrastructure for running the algorithm

Data preparation

- Information and effect period: T_{inf} and T_{eff},
- average features using aggregators over T_{inf},
- predict that the number of alerts will go over a threshold in T_{eff}
- formalize failure: number of alerts is larger than a threshold

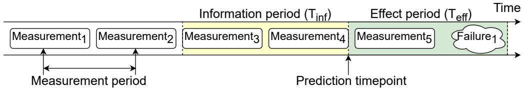


Figure: T_{inf} and T_{eff}

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

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

Algorithm overview

i Optimize data preparation hyperparameters (for some cases can be set from prior knowledge)

ii Build a model to predict the probability of a problem

iii Elicit expert feedback as decision rules"

iv Use ML model and elicited decision rules to predict future problems

Bayesian Optimization of T_{inf} and T_{eff}

We employ BO to find best performance over T_{inf} and T_{eff}. We define a Gaussian process (GP)

prior over functions:

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

In our work, we used the RBF kernel:

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

where γ is a length scale hyperparameter. We use GP as a surrogate for the model's predictive performance.

Acquisition function:

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

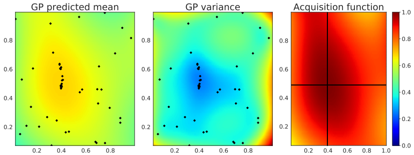


Figure: Bayesian optimization of the information and effect periods.

Human-in-the-loop

Classical ML

i. Humans use predictions, and label the data
ii. Model uses labeled data for training and makes predictions

Our human-in-the-loop implementation

i. Experts provide labels and explain their decision-making process
ii. Model makes predictions and gives intepretable additional information

We aim to build the model so that experts can affect predictions of the model, and the model's decisions are explainable.

Decision rules elicitation

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

Model based on decision rules feedback:

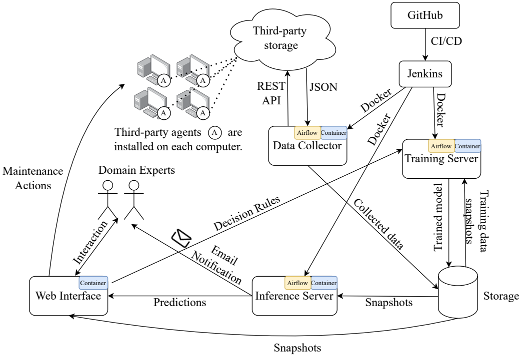
$$C_{fb}(x) = \zeta(\sum_{i=1}^{F} \underbrace{f_i(x)}_{\text{feedback rules}} \underbrace{\mathsf{sim}(X^i_{test}, x)}_{\text{Similarity}} \underbrace{\theta_i}_{\text{Weight}}). \tag{5}$$

The whole model:

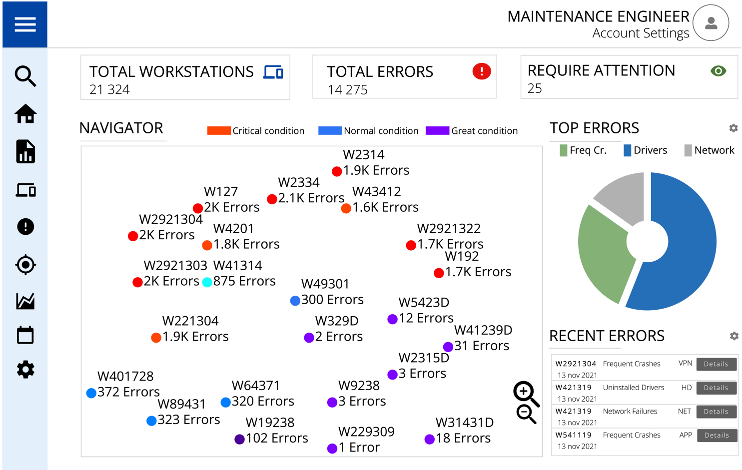
$$C(x) = \underbrace{\alpha}_{\alpha \in [0,1]} \underbrace{M(x, \boldsymbol{\theta}_M)}_{\text{ML model parameterized by } \boldsymbol{\theta}_M} + \underbrace{(1-\alpha)}_{\alpha \in [0,1]} \underbrace{C_{fb}(x, \boldsymbol{\theta}_f)}_{\text{Feedback model}}, \tag{6}$$

Deployment scheme

- 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



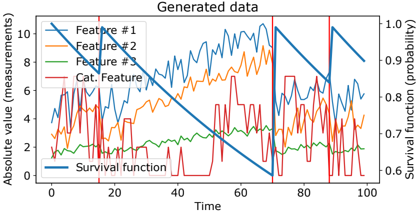
Interface for exploration



The most problematic workstations can be explored via t-SNE embeddings

Experiments

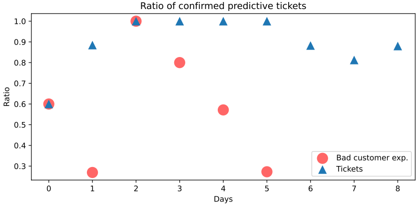
1. Synthetic data



2. Real historical data

Algorithm	Inf. period	Eff. period	f_1-score	precision	recall
Logistic regression	31.9h	119.9h	0.51 ± 0.02	0.88 ± 0.02	0.35 ± 0.01
ExtraTrees classifier	31.9h	119.9h	0.76 ± 0.01	0.72 ± 0.02	$\mathbf{0.82 \pm 0.01}$
Gradient boosting	31.9h	119.9h	0.77 ± 0.01	0.87 ± 0.02	0.69 ± 0.02
DRE (3)	31.9h	119.9h	0.77 ± 0.01	$\mathbf{0.89 \pm 0.02}$	0.67 ± 0.02
DRE (5)	31.9h	119.9h	0.78 ± 0.01	0.87 ± 0.02	0.71 ± 0.02
DRE (15)	31.9h	119.9h	0.80 ± 0.01	0.87 ± 0.02	0.74 ± 0.02
DRE (20)	31.9h	119.9h	$\mathbf{0.81 \pm 0.01}$	0.84 ± 0.02	0.77 ± 0.02

3. Field-testing



Summary

- Human-in-the-loop practical implementation
- The core is decision rule elicitation
- Synthetic, collected, and field experiments

- Future: HITL in other PdM domains or other applications (e.g., healthcare)

- Preprint: `https://arxiv.org/abs/2206.11574`
- Correspondence: `alexander.nikitin@aalto.fi`