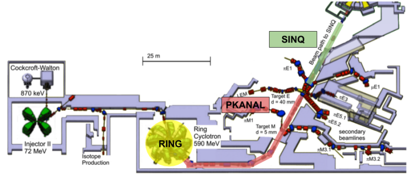
Towards forecasting interlocks of PSI HIPA

Andreas Adelmann Sichen Li Asya Pentina
Davide Reggiani Jochem Snuverink

PAUL SCHERRER INSTITUT

Problem

- interlock system detects anomalies and stops beam operation
- there are false alarms
- idea: forecast interlocks and (try to) prevent them



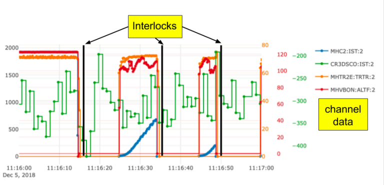


Image credit: Sichen Li

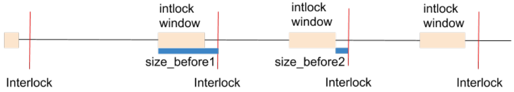
Data

- historical data for 450 sensors across the machine
- discretize time (up to 0.2 seconds)
- window of fixed size (295s) as input features

Window definition

- **"Interlock"** - interlock happens in 5 seconds

window = 5min - 5s = 295s



- **"No interlock"** - no interlock in the near future

window = 5min - 5s = 295s
size_trim = 10min

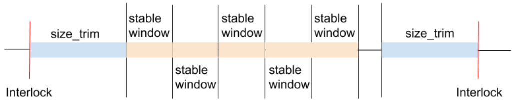


Image credit: Sichen Li

Group Lasso

$$\frac{1}{N}\sum_{i=1}^{N}\log\left(1+\exp\left(-Y^i\sum_{c=1}^{N_c}\sum_{t=1}^{N_t}w_{ct}X_{ct}^i\right)\right)+\lambda\sum_{c=1}^{N_c}\sqrt{\sum_{t=1}^{N_t}w_{ct}^2}$$

- N - total number of training examples (2366)
- N_t - number of timestamps (1476)
- N_c - number of channels (450)

Group Lasso

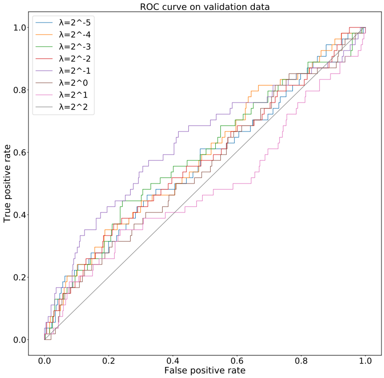
$$\frac{1}{N}\sum_{i=1}^{N}\log\left(1+\exp\left(-Y^i\sum_{c=1}^{N_c}\sum_{t=1}^{N_t}w_{ct}X_{ct}^i\right)\right)+\lambda\sum_{c=1}^{N_c}\sqrt{\sum_{t=1}^{N_t}w_{ct}^2}$$

- N - total number of training examples (2366)
- N_t - number of timestamps (1476)
- N_c - number of channels (450)

depending on λ only a few channels will be used

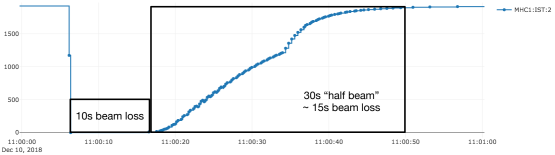
First results

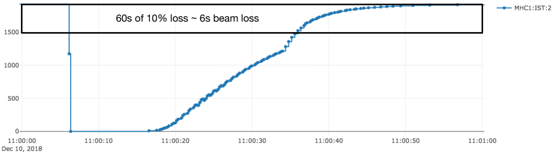
interlocks are rare (10% in the data, even less in reality) $\Rightarrow$ accuracy is not the right measure of success



no clear leader

Loss function definition





Model \ True	Interlock	Stable
Interlock	6	6
Stable	25	0

Image credit: Jochem Snuverink

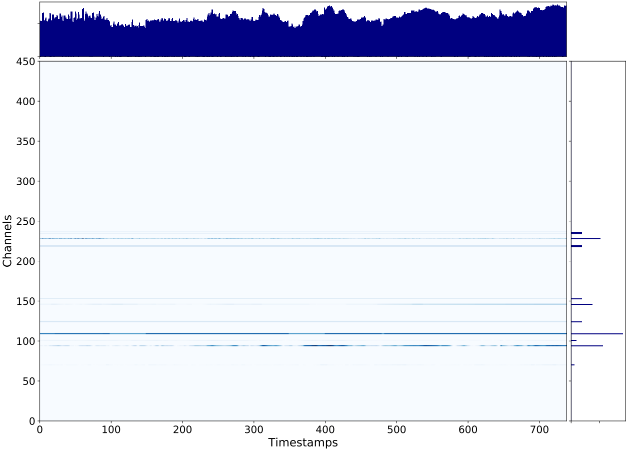
More results: performance

Procedure: select the regularization constant and the threshold on the validation set, apply to the test set

Method	Loss	Savings
Baseline	2.69	
All window	2.56	4.6%
Half window	2.46	8.5%
Quarter window	2.47	8%

- some improvement over not capturing any interlocks
- original window length too long (overfitting?)
- what if we move further away from the interlock?

More results: selected channels



Outlook

- switch to cross-validation
- use the defined loss function in the formulation of the training objective
- look at the effect of moving the window further away from/closer to interlocks
- joint regularization over time and channels?
- move to non-linear models (LSTM?)