Machine learning for failure prediction framework

Christoph Obermair, TE-MPE-PE
Thomas Cartier-Michaud, TE-MPE-PE
Andrea Apollonio, TE-MPE-PE
Lee Millar, CLIC BE-RF
Franz Pernkopf, TU-Graz
Presentation Outline

Presentation of preliminary analysis results:

1. Overview of Data-Driven Models
2. CLIC Analysis Workflow
3. Generic Analysis Workflow
1. Overview of Data-Driven Models

Data-Driven Models

- Manual Feature Engineering
  - Manual Threshold
  - Probabilistic
  - Classic Machine Learning
- Automatic Feature Engineering
  - Discriminative Deep Learning
  - Generative Deep Learning

Physical Models

Approach 1:
Use physics knowledge to define model.

Requirement:
Deep understanding of given use case

Approach 2:
Use model to gain physics knowledge.

Requirement:
Data sets to measure performance e.g.: \( D = \{(x_1, y_1), \ldots, (x_N, y_N)\} \)

Content from M. Maciejewski and H. Fawaz et al., „Deep learning for time series classification“, 2019
## 2. CLIC Analysis Workflow

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Exploration</th>
<th>Modeling</th>
<th>Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>Clean Data</td>
<td>Explore Data Properties</td>
<td>Train &amp; validate Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Progress

<table>
<thead>
<tr>
<th>Supervised Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only with trend data</td>
</tr>
<tr>
<td>Only with event data</td>
</tr>
</tbody>
</table>

- Classification of breakdowns
- Prediction of breakdowns in 20ms
- \(\text{(time synchronized?)}\)
- \(\times\) breakdowns, \(\checkmark\) follow up breakdowns
3. Generic Analysis Workflow

**Workflow:**
1. Transformation: data preprocessing specific to use case → only domain specific part of framework
2. Exploration: data preprocessing specific to models, validation of preprocessing, automated data exploration
3. Modeling: training of classification/prediction models → goal: high accuracy, low complexity
4. Execution: implementation of model into environment, online training and monitoring

settings: parameters, environment, paths
4. Where do we need help?

- Revision: **coding quality, continuous integration**, git usage
- Design choices: general format for storing data (see presentation Lorenz)
- Help:
  - Setting up virtual environment usable for SWAN and Pycharm
  - Set up meaningful tests

The plan is to share this framework with organizations outside cern (TU-Graz, ITER, Melbourne).

https://gitlab.cern.ch/acc-co/mpesoft/clicmlframework.git
5. Comments

- We need guidelines, what is the code scope (e.g. availsim 8 years)
- 2 levels of review:
  - Deep down code reviews
  - More generic reviews
- 1st todo: define conventions & verification progress
  - Meet with tibau
- Meet with daniel to discuss long term
- List in detail what we miss and what we have -> They can add to this list
- Add to Mattermost! Tag when