Cognitive Architecture for Process industries

Peter Bednár (Technical University of Košice, Slovakia)

peter.bednar@tuke.sk
Cross-sectorial Data Lab

Sector A (e.g. aluminium)
- Real-time Plant 1 Operational Platform
- Predictive Functions
- Operational Data

Sector B (e.g. plastic)
- Real-time Plant 2 Operational Platform

Sector C
- Real-time Plant 3 Operational Platform
- Real-time Plant 4 Operational Platform

Cross-Sectorial Data Lab Platform

<table>
<thead>
<tr>
<th>Site Environment</th>
<th>ERP</th>
<th>MES</th>
<th>SCADA</th>
<th>PLCs</th>
<th>Sensors/Actuators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prediction signals

Real-time Plant Operational Platform

Operational Data

Predictive Functions (from Data Lab Platform)

Operational Data (to Data Lab Platform)
Cross-sectorial Data Lab

- Model repository
- Semantic framework
- Development tools
- Distributed Data Processing Framework
- Distributed database
- Batch processing
- Realtime processing
- Distributed file system

- Messaging service
- Data replication service

models

real-time data

batch data
Semantic tools

- **Knowledge sharing**
  - Problem understanding
  - Data understanding
  - Validation

- **Coding**
  - Notebooks combining code, documentation and data/model visualization
  - Repetitive tasks – constrained by data
Semantic tools

- Problem and Data understanding
  - Domain problems -> Data mining tasks
  - Data elements
  - KPIs
- Data pre-processing, modelling, model validation
  - Automatic code generation
Problem understanding

• **Narrative description**

• **Inconsistent formulation of the problem**
  • Highly depends on the experience of the data scientist
  • Consistency about 80-85% for task definition, inputs, outputs and KPIs

• Can we use LLMs to extract task definition from the textual description of the problem?
**problem:** The anodes used in aluminum smelting process are made by mixing petroleum coke with coal tar pitch (binder) to form a paste. Recycled anode butts are also used as filler aggregates and added to the mixture of coke and pitch. Anode is then formed to the compact block using the vibration process and then baked at high temperature around 1200°C. ...

**task:** classification

**input data:** material ratio, time course of vibration, ...

**output data:** predicted anode quality

**KPIs:** material/energy savings, emissions, ...
Testing problems

example problem 1
example problem 2
example problem 3
example problem 4

target problem

task: ?
input data: ?
output data: ?
KPIs: ?

- 5 problems from various domains (industry, economy, astronomy and medicine)
- 50 variants for each description generated by Open AI gpt-4
Preliminary results

- Precision >95
- Average cross-domain recall 87
  - Problematic input data and domain KPIs
- Average intra-domain recall 87.4
  - Medical domain
- More use cases needed for validation
eSAAM 2023
on Cloud to Edge Continuum

Thank You