Flexibility of Modular and Accountable MLOps Pipelines for Cyber Physical Systems

IARIA Congress 2022

Philipp Ruf, Christoph Reich

Djaffar Ould-Abdeslam

Contact: philipp.ruf@hs-furtwangen.de

About the Presenter

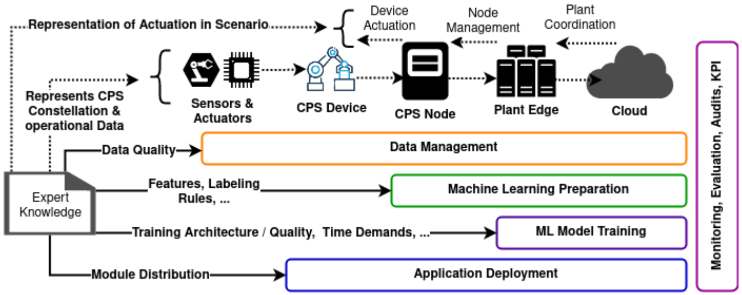
Philipp Ruf is academic staff at the faculty of computer science of the Hochschule Furtwangen University (HFU) and is involved in research projects at the institute for data science, cloud computing and IT-security (IDACUS).
His research topics and areas of interests are cyber-physical systems, infrastructure solutions, security and machine learning.

Agenda

Introduction

- Machine Learning (ML) is applied in the industry by now
  - Solutions for specific problems are often presented in detail
  - Literature often lacks integration and Management steps
  - Respective technology and frameworks impact security policies
- Management of digital objects in Cyber-physical Systems (CPS)
  - Using Digital Twins (DT) of Shopfloor devices
  - Ability of fine-grained modularization and configuration
  - Interconnection of modules forming a task pipeline
  - Implementation of common module interfaces
  - Exchangeable and accountable task-related deployments
- Machine Learning Operations (MLOps)
  - Persistence of data versions whenever data is processed
  - Formalizing modules for the different phases e.g., Data Management, ML preparation, Model training & Deployment within applications
  - Techniques & tools for creating reproducible ML Pipelines e.g., re-runs of a pipeline constellation using a data-version results in same outcome

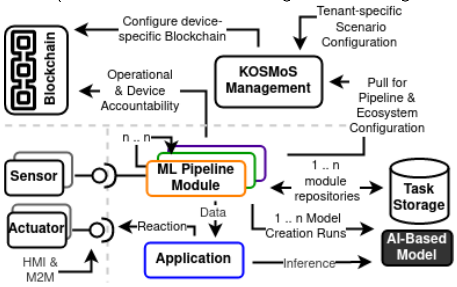
Hierarchical Setups of Cyber-Physical systems


Representation of Actuation in Scenario
Device Actuation
Node Management
Plant Coordination
Represents CPS Constellation & operational Data
Sensors & Actuators
CPS Device
CPS Node
Plant Edge
Cloud
Data Quality
Data Management
Expert Knowledge
Features, Labeling Rules, ...
Machine Learning Preparation
Training Architecture / Quality, Time Demands, ...
ML Model Training
Module Distribution
Application Deployment
Monitoring, Evaluation, Audits, KPI

- Pipelines formalized by Expert Knowledge (JSON Object, XML, ..)
  - Implementation of MLOps-conform modules
  - Module-specific Evaluation, KPIs, Audits, Monitoring, etc.
  - Physical placement of modules in relation to CPS constellation and the problem on hand
- Technical dept and anti-patterns must be considered
- Various frameworks for defining, composing and deploying modules

Accountable Module Deployment Aspects

Research Project *KOSMoS* (Collaborative Smart Contracting Platform for digital value-added Networks)



Blockchain
Configure device-specific Blockchain
Tenant-specific Scenario Configuration
Operational & Device Accountability
KOSMoS Management
Pull for Pipeline & Ecosystem Configuration
n .. n
Sensor
ML Pipeline Module
1 .. n module repositories
Task Storage
Actuator
Data
Reaction
1 .. n Model Creation Runs
AI-Based Model
HMI & M2M
Application
Inference

- Enable and deploy containerized cross-company scenarios
- Focus on generic shopfloor environments
  - Global Ecosystem enables definition of client-side operations

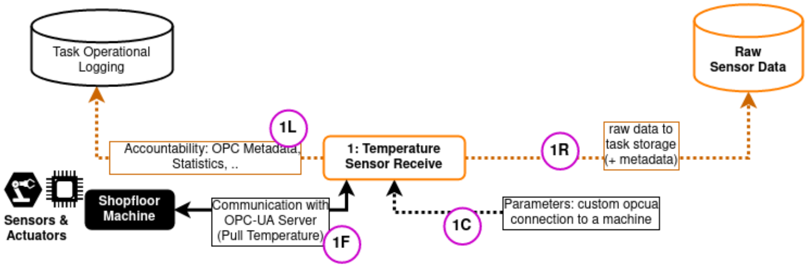
    Blockchain, Identity Provider, Container Registry, Secret Storage, ..
  - Dynamic deployment in client-side environment
  - Trust among participants and accountability of each operation by blockchain technology
- Additional decoupled processing logic by *Smart Contracts*
- Extensible framework enables a variety of scenarios

Modular and Flexible Pipelines

- Modules apply common interfaces (as indicated by the circles the following slides)
  - **C**onfiguration e.g., parameters used by the module logic
  - **F**oreign System Interaction e.g., Communication with systems foreign to the module, e.g., shopfloor devices, other modules, etc.
  - **L**ogging e.g., persist statistics, metadata and additional information in order to increase accountability
  - **R**esults e.g., storing the modules outcomes
  - **i**nput e.g., using the data previously persisted by a foreign module
- Each module persists data in its task-specific storage

  e.g., decoupling & allowing parallel access
- MLOps scenario with one sensor (following slides)
  - Receive Temperature values e.g., realize a connection with the physical device
  - Ensure data quality e.g., assess the basic quality of a dataset version
  - Perform preprocessing e.g., labeling of dataset version intervals
  - Train models e.g., predict a label for the next interval

Modular and Flexible Pipelines

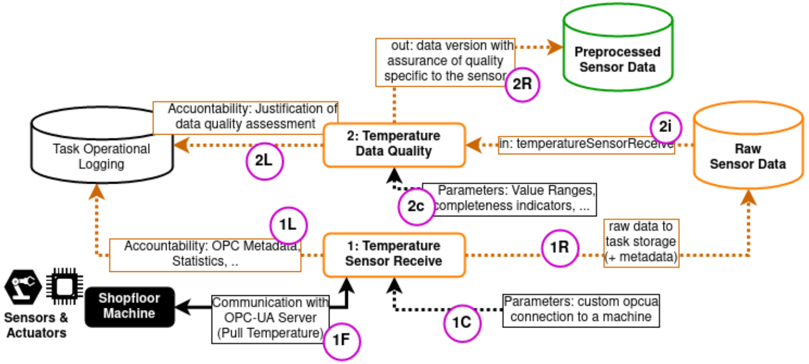


Task Operational Logging
Raw Sensor Data
1L
Accountability: OPC Metadata, Statistics, ..
1: Temperature Sensor Receive
1R
raw data to task storage (+ metadata)
Sensors & Actuators
Shopfloor Machine
Communication with OPC-UA Server (Pull Temperature)
1F
1C
Parameters: custom opcua connection to a machine

Receive sensor values from physical device

- Configuration with device protocols e.g., Open Platform Communications Unified Architecture (OPC/UA) and target interface
- Persistence of sensor values alongside additional metadata
- Logging of additional module-specific statistics

Modular and Flexible Pipelines

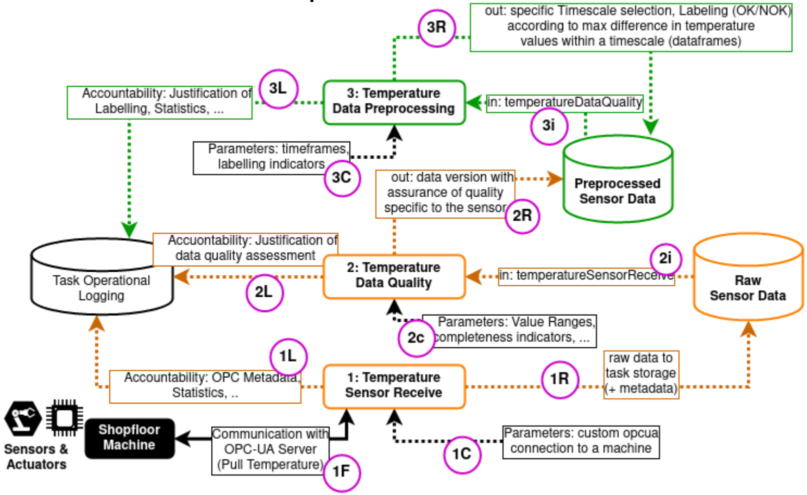


out: data version with assurance of quality specific to the sensor
2R
Preprocessed Sensor Data
Accountability: Justification of data quality assessment
Task Operational Logging
2L
2: Temperature Data Quality
2i
in: temperatureSensorReceive
Raw Sensor Data
Parameters: Value Ranges, completeness indicators, ...
2c
1L
Accountability: OPC Metadata, Statistics, ..
1: Temperature Sensor Receive
1R
raw data to task storage (+ metadata)
Sensors & Actuators
Shopfloor Machine
Communication with OPC-UA Server (Pull Temperature)
1F
1C
Parameters: custom opcua connection to a machine

Ensure basic data quality

- Configure value ranges & completeness indicators
- Receive previously persisted data
- Persistence of assessed data
- Logging of data quality assessment justification

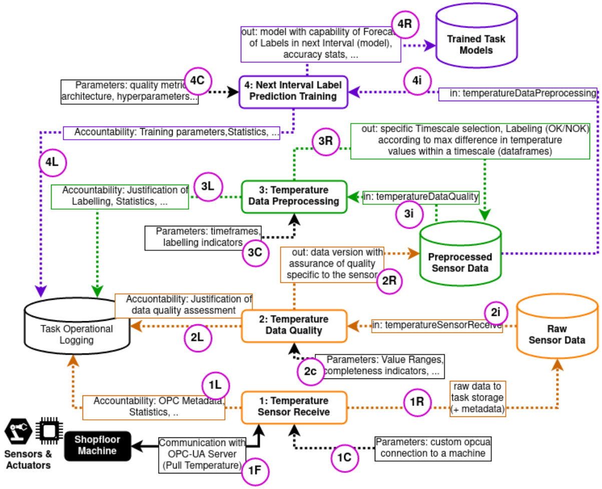
Modular and Flexible Pipelines


3R
out: specific Timescale selection, Labeling (OK/NOK) according to max difference in temperature values within a timescale (dataframes)
Accountability: Justification of Labelling, Statistics, ...
3L
3: Temperature Data Preprocessing
in: temperatureDataQuality
3i
Parameters: timeframes, labelling indicators
3C
out: data version with assurance of quality specific to the sensor
2R
Preprocessed Sensor Data
Accuontability: Justification of data quality assessment
Task Operational Logging
2: Temperature Data Quality
in: temperatureSensorReceive
2i
Raw Sensor Data
2L
Parameters: Value Ranges, completeness indicators, ...
2c
1L
Accountability: OPC Metadata, Statistics, ..
1: Temperature Sensor Receive
1R
raw data to task storage (+ metadata)
Sensors & Actuators
Shopfloor Machine
Communication with OPC-UA Server (Pull Temperature)
1F
1C
Parameters: custom opcua connection to a machine

Preprocessing of previously gathered temperature data

- Automatable data cleansing technique configurations
- Persist labeled data set version, ready for ML training
- Logging of labeling processing

Modular and Flexible Pipelines


out: model with capability of Forecast of Labels in next Interval (model), accuracy stats, ...
4R
Trained Task Models
Parameters: quality metric architecture, hyperparameters...
4C
4: Next Interval Label Prediction Training
4i
in: temperatureDataPreprocessing
Accountability: Training parameters,Statistics, ...
4L
3R
out: specific Timescale selection, Labeling (OK/NOK) according to max difference in temperature values within a timescale (dataframes)
Accountability: Justification of Labelling, Statistics, ...
3L
3: Temperature Data Preprocessing
in: temperatureDataQuality
3i
Parameters: timeframes, labelling indicators
3C
out: data version with assurance of quality specific to the sensor
2R
Preprocessed Sensor Data
Accuontability: Justification of data quality assessment
Task Operational Logging
2L
2: Temperature Data Quality
in: temperatureSensorReceive
2i
Raw Sensor Data
Parameters: Value Ranges, completeness indicators, ...
2c
1L
Accountability: OPC Metadata, Statistics, ..
1: Temperature Sensor Receive
1R
raw data to task storage (+ metadata)
Sensors & Actuators
Shopfloor Machine
Communication with OPC-UA Server (Pull Temperature)
1F
1C
Parameters: custom opcua connection to a machine

ML model training

- Configure the ML model training and other module-specific parameters e.g., architecture, hyperparameters, quality metrics, etc
- Persist model alongside evaluation & module statistics

Dynamic and Extensible Pipelines

- Deploying resulting models e.g., dedicated REST API, integration into application, microservices, etc. can be part of CI/CD
- Obvious increase in pipeline complexity when additional sensor(s) become available in addition, Deadlocks or Bottlenecks can occur
- Utilization of multiple modules in parallel in different locations
- Exchangeable modules 'on-the-fly' & extensible pipelines
- Constant monitoring and evaluation of each module, technology, tool and system involved in the scenario

Conclusion & Future Work

- Flexible, module-based and accountability-enhanced pipeline
  - Selected quality requirements on module interactions
  - Dynamic alterations to existing module pipelines & module configuration
  - Example scenario with one data source & discussion of increased complexity when the shopfloor changes
- Future Work
  - Investigate the integration of AutoML techniques
  - Using simulations of industrial environments for testing scenario specific pipeline compositions beforehand

Thank you for the Attention