CaRL

Causal Reinforcement Learning and Visual Reasoning for Transparent AI

Niccolò Filipas[1], Roman Kern[2] , Marc Streit[3] ,Belgin Mutlu [1]
Pro2Future GmbH[1,] TUG-IML[2], JKU-ICG[3]
[1] Sandgasse 34, 8010 Graz, Austria
[2] Inffeldgasse 16b/I, 8010, Graz, Austria
[3] Altenberger Straße 69, 4040 Linz, Austria

MOTIVATION & GOALS

Industrial faults require more than classification: useful decision support should expose how abnormal behavior propagates through a process. CaRL tests whether a **known causal graph** can guide transparent anomaly detection in **multivariate time series**. The current goal is to learn normal process dynamics with lagged structural equations, detect faults from deviations between observed and causally expected behavior, and **recover** the **healthy process** state using **reinforcement learning**.

Project FactBox

Project Name	CaRL
Project ID	StratP II A.S-2
Duration	38 Months

Area 3
Area Analytics

Project Lead
DI Dr. Belgin Mutlu

APPROACH

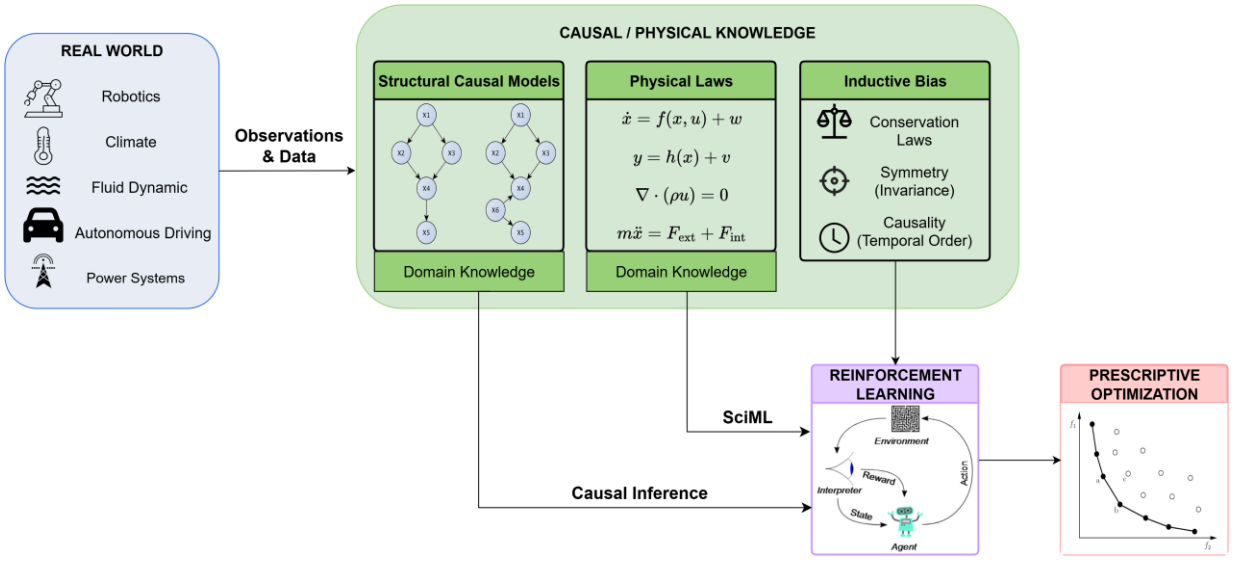
- Preprocess multivariate process data and leverage a known **causal graph** to capture relationships between variables.
- Learn normal process dynamics and estimate expected system behavior from **causal** and **temporal dependencies**.
- Detect faults by measuring deviations from expected behavior and applying statistical anomaly thresholds.
- Support **transparent diagnosis** by tracing anomalies through the causal structure of the process.
- Generate **counterfactual explanations** to identify potential causes and corrective interventions.
- Evaluate **counterfactual** recovery strategies by simulating alternative process states.
- Use **Reinforcement Learning** (RL) to recommend actions that restore normal operation with minimal disruption.

CONTRIBUTION

Scientific contribution
- Graph-guided dynamic modeling of normal industrial process behavior using process-structure knowledge.
- Interpretable variable-level anomaly detection through residuals linked to known process relationships.
- Run-based evaluation framework assessing anomaly occurrence, detection reliability, and time-to-detection.
- Improved understanding of process deviations through causally informed diagnostics rather than black-box anomaly scores.

Practical contribution
- Reusable end-to-end pipeline for training, storing, and applying structural models.
- Support for early anomaly detection and operational monitoring.
- Causally informed diagnostics for understanding process deviations.
- Foundation for future counterfactual recovery and process optimization applications.

FRAMEWORK



Contact: Belgin Mutlu, Pro2Future GmbH, belgin.mutlu@pro2future.at, +43 664 88371323.
Acknowledgement: This work was supported by Pro²Future II (FFG, 911655) and JKU, Institute of Pervasive Computing