Ensuring physical coherence in data-based models

Industrial digitalization together with the improved data storage and computation capacities facilitate the creation of data-based models of the processes (e.g., via deep learning). These models are to be used for prediction, sometimes asked to provide a response at operating conditions different from the ones used for training such models. As machine-learning methods usually rely on the data quality and quantity, this may involve serious risks when building black-box models for industrial process systems that usually operate around the same steady states, where on-site experimentation is very expensive (poor diversity), or measurements are corrupted by outliers or sensor noise (poor reliability).

All available information on the process needs to be present in the model building phase. A grey-box model includes both sources of information, plant observation and well-known process physics. Moreover, models not only must match the recorded plant data, but also provide coherent physical responses in other situations of interest.

Methodology

SOS framework

Sum-Of-Squares Constrained Regression

- SOS polynomials are proven to be always positive.
- Local positivity can be ensured via the Positivstellensatz theorem.
- SOS programming is convex optimisation (semidefinite prog.)

The SOS approach to machine learning:

\[
\begin{align*}
\min_\beta & \sum_{i=1}^N \sigma_i(y_i - p(\beta, x_i))^2 \\
\text{s.t.} & \quad c(\beta, x) \geq 0, \quad x'g(x) > 0 \\
\beta & \leq \beta \leq \bar{\beta}
\end{align*}
\]

Polynomial candidate model in \( x \)

Derivatives of \( p \) w.r.t. \( x \) are also polynomials in \( x \)

It allows enforcing constraints on:
- bounds (zero-order derivatives)
- slopes (1st derivatives)
- curvatures (2nd derivatives)

Local region defined by a polynomial boundary \( g \)

A first-principles skeleton combined with data reconciliation provides estimations coherent with the known physics of the process.

Constrained regression allows including any further physical insight that the modeler/engineer may have.

The SOS framework has advantages over the “blind” regularisation, especially when visual inspection is impossible (e.g., when the no. of input variables is larger than 2).

A well-behaved response for inter/extrapolation is preferred over a perfect fit: data may be corrupted by noise, or many samples contain the same information.

A two-stage systematic procedure

1\(^{st}\) Stage inputs:
- Process data, Plant historian, experiments...
- Process physics, Well-known first-principles equations.

1\(^{st}\) Stage outcome:
- Coherent estimations over time for all variables present in the model equations.

2\(^{nd}\) Stage inputs:
- Virtual data, Corrected actual + estimations coherent with the process physics.
- Desired features to enforce in the data-based model response.

2\(^{nd}\) Stage outcome:
- Well-behaved black-box equations.

Final result:
- Reliable grey-box model for prediction.

Example: modelling the heat-transfer coefficient in a heat exchanger

Model unreliable for extrapolation in the reddish zone.

Physically coherent model in the whole region of operation.

Want to know more?

Developers & contacts:

Dr. José Luis Pitarch – josé.pitarch@autom.uva.es
Prof. Dr. César de Prada – prada@autom.uva.es

www.copro-project.eu

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 723575