Machine Learning for Digital Twins in Virtual Materials Design

Roland Aydin
Ingredients of a Digital Twin

- Theory/Model specification
- Model implementation
- Integration / Validation
- Data
Ingredients

1. **Theory/Model specification**
   - noise
   - simplifications
   - confounders

2. **Model implementation**
   - simulation code
   - multiscale

3. **Integration**
   - data driven validation

4. **Machine Learning**
   - data from adjacent domains
   - data with high dimensionality

---

[Image of simulation code]

[Image of multiscale model]

[Image of data driven validation]
Showcase 1: Machine Learning for optimizing polymer membranes

Motivation

• Machine Learning for functional dependencies (objective functions) which cannot be quantified easily

• Correlate process parameters and the resulting membrane characteristics
  Invert to elicit the optimal process parameters for the intended membrane properties
Showcase 1: Machine Learning for optimizing polymer membranes

- What if the *optimal* targets are unknown? (process parameters → unknown_toplogy_characteristics)
- What if manual labeling is prohibitively expensive?
Showcase 2: Machine Learning for Mg-Corrosion

Motivation

- Feature selection:
  - to enable surrogate models (by improving the signal/noise ratio)
  - to drive experimental design choices

- Data-driven dimensionality reduction:
  using Machine Learning as an intermediate stage to bootstrap analytical understanding

E. Schiessler, T. Würger, S. V. Lamaka, R. H. Meißner, C. Cyron, M. L. Zheludkevich, C. Feiler
Showcase 2: Machine Learning for Mg-Corrosion

Molecular Descriptors
- Constitutional
- Connectivity
- Topological
- Ring

Categories: e.g. LUMO

Magnesium Dissolution Modulators

Feature Selection
- Pick top $n$ features
- Pick top $n$-tuple

Deep Learning Model
- Mor04m
- E1p
- LUMO
- P_VSA_MR5
- Mor22s

Selected Molecular Descriptors

Approach I: ANOVA
Approach II: RFE

Corrosion Inhibition Efficiency for ZE41 (IE/%)
Showcase 3: Synthetic data in electron microscopy

Motivation

• Use case:
  Machine Learning to solve the shine-through effect in FIB data of nanoporous materials

• Use case:
  Machine Learning to correct original data (FIB slice allocation)
Showcase 3: Synthetic data in electron microscopy

Results

In case of insufficient data to train the Machine Learning model:
- construct our own synthetic data in 2 steps:
  - virtual microstructures → MC simulations → synthetic data usable for supervised learning

Machine learning for automatic segmentation

(if sufficient quantities of data are available)
Showcase 3: Synthetic data in electron microscopy
Results: Improved segmentation, Correcting original data

Trushal Sardhara & Martin Ritter & Alexander Shkurmanov & Yong Li & Shan Shi (TUHH), Christian Cyron (Hereon)
Showcase 4: Neural Architecture Search: the „Surgeon“ Motivation

- Machine Learning models have a large number of hyperparameters (learning rate, activation function, gd optimizer, batch size, topology choices, etc.)

- Finding the best combination for a given problem is infeasible (combinatorial explosion)

- Network choices have to reflect a deep understanding about the problem domain that is not necessarily present (CNN vs. FFNN, LSTM vs. Transformers, best type of autoencoder, etc.)
Modification choices can also e.g. uphold physical constraints (e.g. Navier-Stokes, PINN alternative); optimize runtime constraints (#weights)
Take-Home Message

• Digital Twin: a *toolkit* rather than a monolithic sibling

• The role(s) of Machine Learning for Digital Twins in VMD
  • Bridging the gaps between the digital twin components
  • (In some cases)
    Building understanding, transfer learning for optimizing data acquisition
  • (In rare cases)
    Surrogate models