Bio

Perman Jorayev is a final year PhD student in the department of Chemical Engineering and Biotechnology at Cambridge University. His research focuses on utilising machine learning (ML) algorithms and robotics automation to accelerate process development in the pharmaceutical industry. Perman holds a BSc in Chemistry and Chemical & Biomolecular Engineering at Korea Advanced Institute of Science and Technology (KAIST), where he has worked on developing fluorescent molecules for nerve agent detection and synthesising porous materials for gas capture and storage. As a visiting researcher, Perman has also worked on live cell imaging of focal adhesions and crude-oil upgrading projects at EPFL, Switzerland and MIT, USA, respectively.

Machine-assisted holistic workflow to robust process development in the pharmaceutical industry

Perman Jorayev,[a,b] Simon Sung,[b] Mohammed Jeraal,[b] Alexei A Lapkin[a,b]

[a] *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, United Kingdom*
[b] *Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd. 1 CREATE Way, CREATE Tower #05-05, Singapore 138602, Singapore*

Process development of novel transformations is often a laborious and complex task. This is mostly due to the difficulties in identifying the underlying chemical and physical parameters that affect the process objective(s), quantifying the nonlinear interactions between them, and lack of data.

When optimising for continuous variables only, conventional and Bayesian optimisation-based DoE have been demonstrated. However, when expanded into discrete variables including solvents, reagents, or ligands, these approaches often struggle with the curse of dimensionality and inefficiencies (in predictive accuracy) of black-box optimisation algorithms.

A common approach to overcome this problem has been demonstrated using molecular descriptors to map the discrete variables onto a continuous space. However, existing approaches often: (1) utilise large number of variables / dimensions (15+) to parameterise solvents, (2) start the optimisation from scratch (i.e., choice of descriptors are not validated to be the most relevant), and (3) do not demonstrate a holistic approach (i.e., keep other reaction variables fixed).

Parameterisation of discrete variables with relevant descriptors could increase the predictive accuracy of surrogate models to unlock new mechanistic insights and to enhance robust and green process development. Demonstrated on three case studies, we address these challenges using a generalisable workflow, by (1) extraction of relevant literature data, (2) identifying and quantifying relevant reaction parameters using supervised ML to generate a priori model, (3) an efficient DoE for sampling from the discrete space using unsupervised ML. Using an (4) automated flow setup, this approach is then integrated into a (5) Bayesian optimisation algorithm to demonstrate a holistic workflow on optimisation of discrete and continuous variables for multiple objectives on a pharmaceutically relevant photoredox amine synthesis.