FOPAM 2026
Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning
Decatur, USA, July 26-30 (2026)

Topology-Guided Response Surface Characterization for ML Model Selection

Shenbageshwaran Rajendiran

Auburn University, Department of Chemical Engineering

ABSTRACT

Machine learning (ML) models are widely used as surrogates for complex optimization problems in process systems engineering; however, selecting an appropriate surrogate remains challenging. Current selection approaches often rely on cross-validation, prior experience, or statistical and gradient-based response-surface descriptors. We hypothesize that topological descriptors such as the Euler characteristic curve (ECC), which captures structural changes across thresholds, provide complementary information for structure-aware surrogate model selection. We evaluated this hypothesis using 41 two-dimensional optimization test functions annotated with four landscape characteristics: modality, ruggedness, abruptness, and geometric complexity. For each function, ECCs were first computed on a dense 25*25 (625 samples) grid using sublevel filtration. Summary metrics from ECC, including the number of ECC jumps, cumulative ECC variation, and maximum ECC jump, were extracted from each curve. Thresholds for classifying each landscape characteristic were calibrated from the dense grid using Youden's index. To evaluate performance under sparse sampling, each function was randomly sampled at 10, 25, 50, 100, 300, and 600 points, with 30 replicates per sample size. The ECC computation was performed using a grid-based response surface representation. Therefore, the randomly sampled values were constructed onto 25*25 grids. Since random sampling does not cover every grid cell, the remaining empty cells were filled using nearest-neighbor cell filling, where each empty grid cell was assigned the value of its nearest occupied cell. ECC summary metrics were then computed for each reconstructed grid, and the dense-grid-derived thresholds were applied to predict landscape characteristics. Prediction performance was evaluated using accuracy and the Matthews correlation coefficient (MCC). Across random-sampling experiments, ECC-based characterization improved with sample size and stabilized at approximately 100 samples. At this sample size, ECC metrics identified landscape characteristics with approximately 0.8 accuracy and 0.8 MCC, compared with approximately 0.9 for both metrics on the dense grid. These results suggest that ECC metrics capture response-surface structure that remains informative under sparse sampling and provide a promising avenue for structure-aware ML model selection. Future work includes extending the current framework to evaluate different sampling techniques, extending to higher dimensions, and relating ECC metrics to the performance of surrogate models with varying levels of complexity.