

Materials Informatics for Coatings Formulations:
**Applied Machine Learning Strategies for
Rapid Reformulation**



Applied Machine Learning Strategies for Rapid Reformulation

Abstract

To meet ever-shifting market, customer and supply demands, leaders in the coatings industry have adopted materials informatics (MI) to accelerate formulation development. Historically, intuition-based knowledge has been a key element of R&D in new coatings development. Through MI-driven product development, historical data alongside expert domain knowledge is captured, preserved, and deployed in the service of accelerated new product development. Below, we discuss the core concepts enabling the successful deployment of MI to coatings problems. This example shows the application of materials informatics to address a common coatings reformulation problem: removing APEO-containing surfactants to achieve regulatory compliance.

Application of Materials Informatics to Coatings Problems

The sequential learning workflow is the core methodology for the application of materials informatics, and represents an active learning approach wherein existing data is leveraged alongside expert domain knowledge to propose experimental candidates with a high likelihood of achieving technical goals. The outcomes of machine-proposed experiments are fed back into the dataset, enhancing the algorithm's "understanding" of the system's region of interest. An overview of the sequential learning methodology is shown in Figure 1.

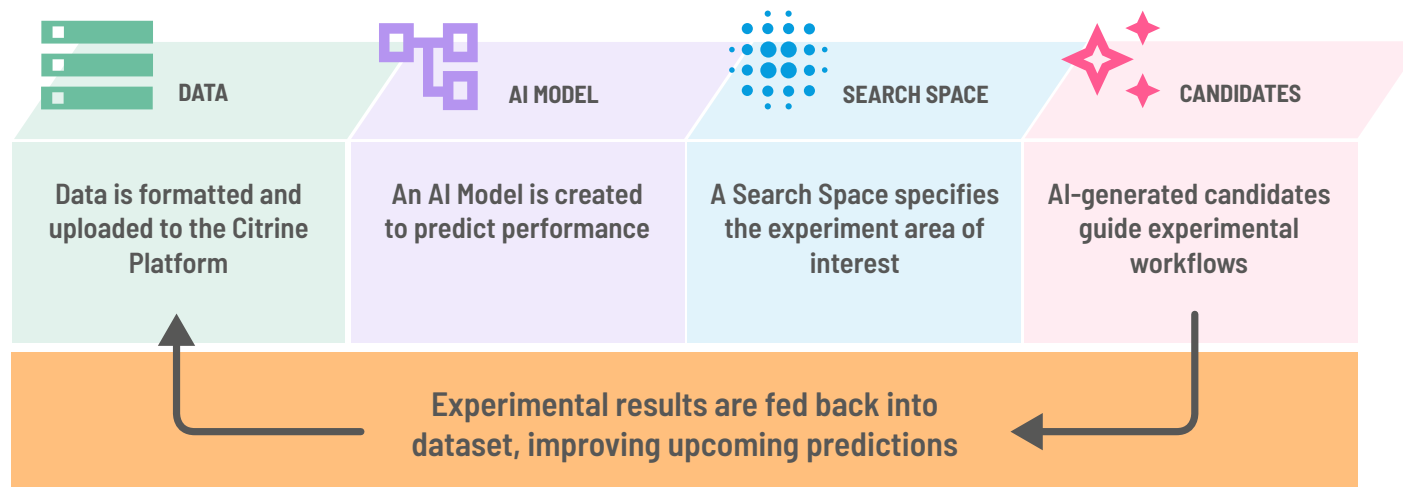


Figure 1: The Sequential Learning Workflow for Coatings R&D

Domain Aware Data Structures. In the coatings industry, data is commonly available in limited capacity, therefore it's critical to maximally leverage all available data. This is accomplished by capturing all relevant information in a material history. The material history captures relevant information about processing conditions and performance properties, such that every necessary detail for experimental reproducibility is recorded. In addition to capturing the history of production, it is useful to capture information about constituent materials or ingredients, such as the HLB value of a surfactant, or the oil absorption of a pigment. By capturing these physical properties of the raw materials, AI models can understand and predict the behavior of a system. In addition to enforcing rigorous record keeping and retention of high quality data, the material history structure enables the automatic creation of no-code machine learning models tailored to your technical problem.

Chemistry Aware Modeling. Experimental data is utilized to build an ML model which predicts the properties of interest in a coatings system. A key advantage of Citrine's graph AI model approach is that it enables automatic feature engineering, as well as custom domain knowledge integration. Citrine's graph AI models build on formatted data to create features which are broadly relevant across coatings formulation problems, these include the amounts and types of ingredients, the properties of constituent components, as well as features describing known chemical components. Citrine graph models also enable the direct injection of domain knowledge via analytical expressions; for instance if the ratio of surfactant to thickener is impactful to the stability of a product, this information may be expressed directly in the model. Once a graph model is built, it is automatically assessed with cross-validation, and the performance may be evaluated, prompting further refinement prior to proceeding to the design phase or at any point during the sequential learning workflow.

Effective Design with Search Spaces. Search spaces represent the virtual lab bench, or "what is possible" to fabricate or formulate. Search spaces may represent any desired breadth in degrees of freedom, but the highest quality search spaces represent a specific hypothesis, for instance: "If I eliminate or replace a specific component of my existing formulation, can I still achieve a high performing product?" Search spaces are defined by specifying the inclusion bounds on specific ingredients, types of ingredients, and associated process parameters.

Efficient Candidate Exploration. The tangible output of an AI driven workflow is the production of candidate recipes, which represent the data-driven and domain-knowledge guided experimental plans, aimed at achieving users' product development goals. This final step in the sequential learning process enables your existing personnel to combine their intuition for system performance with data driven proposals to quickly achieve target behavior. Experimental results are fed back into the dataset to complete the sequential learning loop, effectively enabling the AI system to learn with each new piece of information.

Demonstration: Replacing APEO Based Surfactants

We discuss the AI-driven approach to a common problem in the coatings industry: reformulating an existing product to achieve regulatory compliance. In this scenario, an existing architectural paint product line contains a component which will soon be banned from production. The goal is to reformulate the product such that undesired ingredients are not included, but the performance of the final product remains the same as the existing commercial product.

Challenge:

Formulating new paints or re-formulating existing paints can require many iterations of latex, thickener, and surfactant blends to hit performance targets while maintaining a stable dispersion. This has become increasingly more challenging as domain knowledge in this space is becoming harder to find. How do we utilize the platform to predict stable paints and instill domain knowledge along the way?

Approach:

Data Onboarding: An existing dataset is structured into a machine readable format. This process involves collecting the data into a tabular format, specifying the paints in the dataset in terms of ingredients, mixtures of ingredients, and annotating those ingredients with relevant properties such as the HLB values and molecular structures of surfactants. In addition to the ingredients which are used in the experimental dataset, we include ingredients and properties which may be suitable replacements for the undesired materials. The machine-ready dataset is then uploaded to the Citrine platform to enable ML modeling.

ML Modeling: An automatically generated graph AI model is generated from the structured data. This model accomplishes both feature engineering and ML modeling, taking into account the properties that were specified in the dataset, and enabling the ML model to understand the context of the different ingredients and their properties. Additional domain knowledge is added to the graph model by product developers, who use the point-and-click interface to represent known relationships and system heuristics.

Search Space Development: A search space is developed based on domain expertise and the historical data. Because the goal is to maintain the characteristics of the existing product, we choose to use the same core components which comprise the grind and letdown. Domain expertise dictates that relative identity and amounts of surfactant and thickener have a strong bearing on the solution stability of the product, therefore the search space is configured to enable exploration over a variety of candidate surfactants and thickeners, with the goal of elucidating the optimal components and relative concentrations.

Candidate Generation: Experimental candidates are generated with the explicit aims of achieving identical performance to the existing product, with the simultaneous requirement that no banned substances are included. Candidates are evaluated by product developers, who consider the predicted values and associated uncertainties to choose the optimal experiments to conduct in service of achieving the project goals faster than ever before.

Summary

Product development in the coatings industry is facing increasingly stringent demands. We shared a typical challenge which requires simultaneously increasing performance across many parameters while also removing a growing list of undesirable ingredients. To address these types of needs, more and more industry leaders have adopted the Citrine Platform to accelerate product development outcomes. If you would like to learn more about how to implement scalable, secure, democratized AI inside of your organization from the industry leader, please reach out to us.



citrine.io
info@citrine.io

© 2023 Citrine Informatics Inc. All Rights Reserved.