Rapid New Product Development with Extra Small Datasets using AnneaLT algorithm

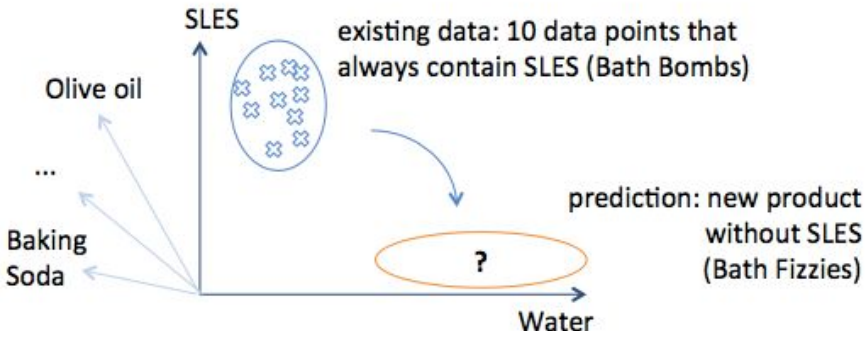
Machine Learning (ML) methods commonly require a large amount of data to fit a model and yield reasonable predictions. However, there are many cases when only limited data are available. Another challenge is prediction of something that was not presented in the historical data.

Find out more about:

- ✓ Prediction provided by AnneaLT compared with existing ML methods for small data amounts
- ✓ Interpretation of the fitted model for better data/process understanding
- ✓ Enhance data-driven fitting by Quantum Computing

Problem Formulation

There is limited data of trials to produce Bath Bombs (hard-packed mixtures of dry ingredients which effervesce when wet. They are used to add essential oils, scent, bubbles or color to bathwater. Usually it contains SLES -a detergent creating foam), one has to determine an optimal proportion of water in Bath Fizzies (which are the same as Bath Bombs but do not form a foam when they effervesce due to the absence of SLES). These two products are produced by the same technological process but have differences in a few components. Namely, all available data consist of 10 points with a non-zero SLES component, while the new product is free of SLES. The problem is to predict a new water concentration for Bath Fizzies based only on 10 points for Bath Bombs. The data in the chart shows the proportions of ingredients at different temperature and humidity when the product is stable. The task is to predict a new value for water content at different temperature and humidity



Data*

Samples / Components	Bath Bombs 1	2	...	10
Baking Soda	0.74	0.83	...	0.65
...	...	...	...	...
Olive oil	0.57	0.77	...	0.60
SLES	0.76	0.60	...	0.46
Water (target)	0.09	0.14	...	0.08

→

Bath Fizzies 1	2	...	10
0.74	0.84	...	0.65
...	...	...	...
0.57	0.77	...	0.60
0	0	0	0
?	**?**	**?**	**?**

Prediction / Samples	Bath Fizzies 1	2	...	10	
Linear Regression	-0.23	-0.12	...	-0.13	negative concentrations
Decision Tree	0.09	0.14	...	0.08	same as training set
Support Vectors	0.11	0.11	...	0.11	constant value
AnneaLT	0.64	0.84	...	0.36	physical prediction

* due to data protection policy all values are presented in dimensionless values without units. Please contact AnneaLT directly for more information.

AnneaLT vs Open-Source Solutions

Although 10 points are not enough to utilize machine learning techniques, one can still fit models and compare the results. Linear Regression is overfits in this case and predicts a nonphysical negative concentration of water. Decision Trees predict the same values as in the training data. Support Vector Regression predicts a constant value close to the mean value of the training set. AnneaLT is the only method that gives the same prediction as would be given by experts (Formulation Chemists and Engineers).

How does it work?

To achieve such results, AnneaLT uses a special mathematical tool for dynamical system modeling. To apply it to the given data set, one has to make two assumptions.

Firstly, we consider production as a continuous process that potentially can be described by a system of differential equations. Instead of deriving the specifics of such equations by experts, AnneaLT implements a neural networks architecture. This allows us to build an AI-based solution and fit the model numerically. Secondly, we consider concentrations of the components in Bath Bombs and Bath Fizzies only as different initial conditions of the dynamical process that describes component mixing.

From a high-level point of view, AnneaLT provides alternate ML models that are able to operate with a small amount of data. Such data-driven models can be used in existing production control pipelines, and provide an optimal prediction based on varying initial conditions of the processes.

How Quantum Computers can enhance the performance?

AnneaLT takes advantage of quantum computing by utilizing commercial Quantum and Digital Annealers. Quantum and digital annealing are two approaches to the development of application-specific hardware devices (annealers) designed for solving binary optimization problems. Such problems are challenging for traditional computing approaches, and the architecture of annealers allows them to solve optimization problems much more efficiently. One approach to utilizing this technology is in the optimal selection of non-zero weights in our Lie transform based neural network. This problem can be converted directly to the type of binary optimization problem that annealers are designed to solve. Another way that AnneaLT uses annealers is to run an efficient optimal parameter search in place of running the model many times.

Conclusion

AnneaLT is a quantum-ready prediction platform that combines machine learning with the ability to use very limited data sets, as well as to predict something that has not been explicitly presented in the historical data. Provided models can be used for determining new optimal conditions of existing processes, commissioning of new machinery, and scaling up production based on the R&D trials.

* due to data protection policy all values are presented in dimensionless values without units. Please contact AnneaLT directly for more information.