Salty Project

High Performance Computing and Machine Learning for Engineering Solvents

Miwako Ito, Wesley Beckner, Yatong Ge, and Jim Pfaendtner

Department of Chemical Engineering, University of Washington, Seattle, WA, 98105, United States

Motivation and Background

Salty Project is a webapp where researchers can:

1. Experiment with predictive computational models.
2. Easily access IL thermophysical data.

ILs (ionic liquids) are excellent solvents due to their low volatility, low vapor pressure, and range of polarities. This has lead to exponential growth in interest for developing new ILs. ILs have applications in batteries and biomass transport.

What is the problem?

NIST (The National Institute of Standards and Technology) provides ILThermo, a database with over 1 million IL thermophysical datapoints. However, accessing the data requires navigating a complicated filter then pasting the results into an Excel document. This is a significant barrier to IL researchers who require thousands of datapoints.

What is the solution?

Salty Project is driven by data visualization. The webapp provides an interactive model in the form of a graphical interface. Users can test a predictive computational model and easily read thousands of datapoints. Additionally, all IL data on the webapp is available in .csv file format.

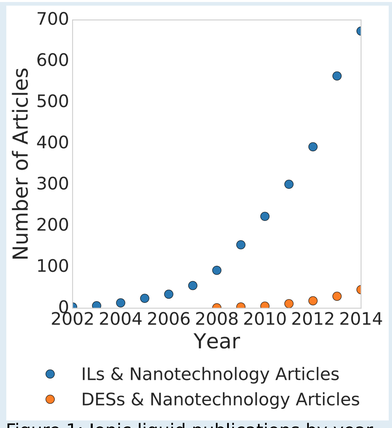


Figure 1: Ionic liquid publications by year.

Abo-Hamad, et al. (2015). Potential applications of deep eutectic solvents in nanotechnology. Chemical Enginering Journal, 273, 551–567. http://doi.org/10.1016/j.cej.2015.03.091 .

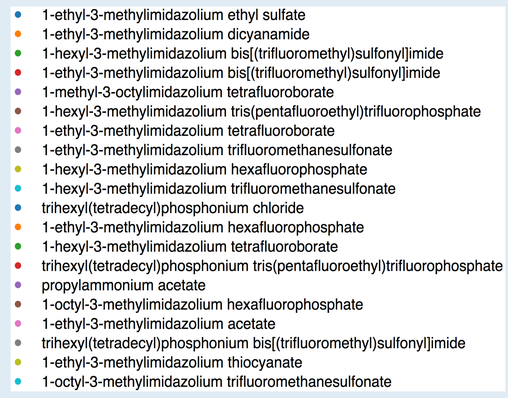


Figure 2: D3 produced legend with names of IL salts.

Webapp Development

The framework of the webapp uses HTML. The frontend (user interface) uses D3. The backend (sever, application, and database) uses Python.

What features does Salty Project include?

D3 is a JavaScript library for data visualization. As the frontend includes the interactive LASSO model, D3 provided the best option for interacting with the IL data and viewing the model output. The interactive LASSO model includes:

1. 9,000 viscosity datapoints.
2. Tooltips for identifying datapoints.
3. A color-coded legend for each type of salt.
4. Panning and zooming functionality.
5. A slider bar for adjusting the alpha value.

With plans to add:

1. Filter-sort interactivity to the legend.
2. Dynamic updates of model improvements.

The current model interacts with static underlying datasets from the Python backend. In the future, creating a dynamic link between the D3 frontend and Python backend will allow the user to train and test their model interactively.

Data Aquirement

31, 326 density datapoints received.

9,000 datapoints for 22 types of salts used in current model.

Feature Generation

Experimental conditions.

Molecular features.

Python-based machine learning package.

Model Training

Used in development of LASSO model.

Open source tool for developing statistical models.

Computational Modeling

Why use high performance computing (HPC)?

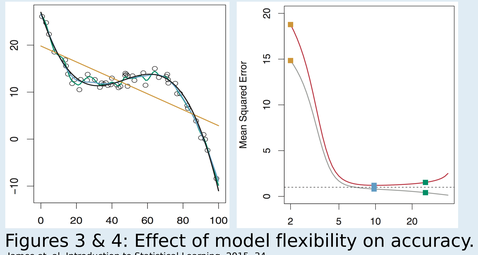
HPC gives researchers a novel method for materials development. Traditional labs undergo exhaustive synthesis and testing when developing new materials. With HPC, researchers can build computational models that predict thermophysical properties of molecules by analyzing empirically obtained data.

Which model does Salty Project use?

Salty Project utilizes a model which predicts the viscosities of ILs. The model, LASSO (Least Absolute Shrinkage and Selection Operator), is a variation of linear regression. Basic linear regression finds the best linear function for fitting a dataset. LASSO improves on the method by adding the hyperparameter alpha. Increasing alpha removes certain features (input) by reducing their coefficient to 0.

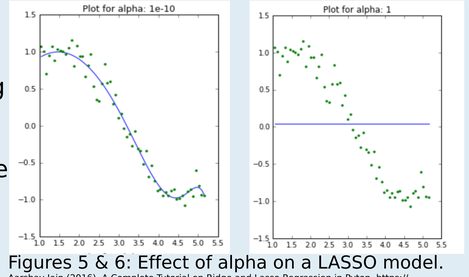
This penalization of features prevents the model from becoming overly flexible and fitting to irrelevant outlier datapoints, a phenomena known as overfitting. Choosing alpha depends on minimizing mean squared error. Features in Salty Project include experimental conditions, and physiochemical descriptors of the cationic and anionic moieties.

Linear function: $y = m_1x + m_2x + ... + m_n x$



Figures 3 & 4: Effect of model flexibility on accuracy.

James et. al. Introduction to Statistical Learning, 2015, 34.



Figures 5 & 6: Effect of alpha on a LASSO model.

Aarshay Jain (2016). A Complete Tutorial on Ridge and Lasso Regression in Pyton, https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-ridge-lasso-regression-python/.

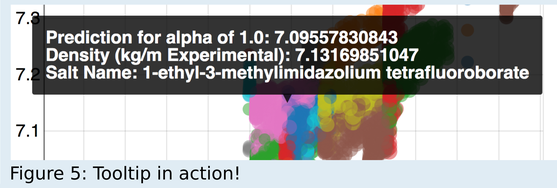


Figure 5: Tooltip in action!

Results

Extra Details

The Salty Project webapp is available for download through GitHub.

An included slider bar allows users to adjust the alpha value and watch the LASSO model change.

Tooltips update as the value of alpha changes.

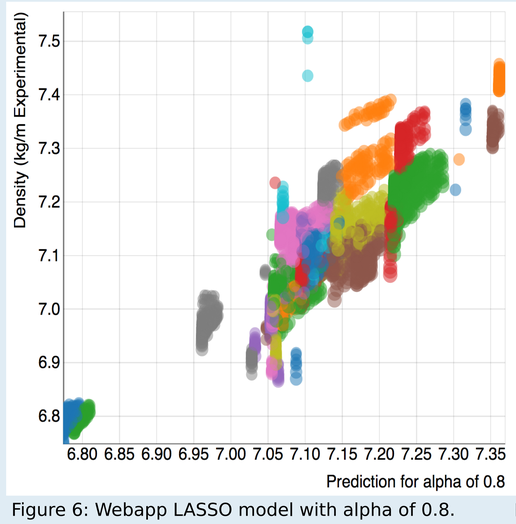


Figure 6: Webapp LASSO model with alpha of 0.8.

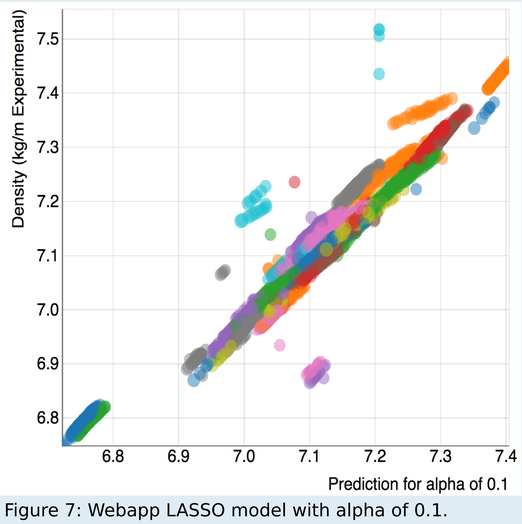


Figure 7: Webapp LASSO model with alpha of 0.1.