Machine Learning Methods in Combustion Chemistry: Mechanism Reduction and Reaction Network Analysis

Zirui Liu[1], Dake Li[2]

PRINCETON UNIVERSITY

[1] Department of Mechanical & Aerospace Engineering, Princeton University
[2] Department of Economics, Princeton University

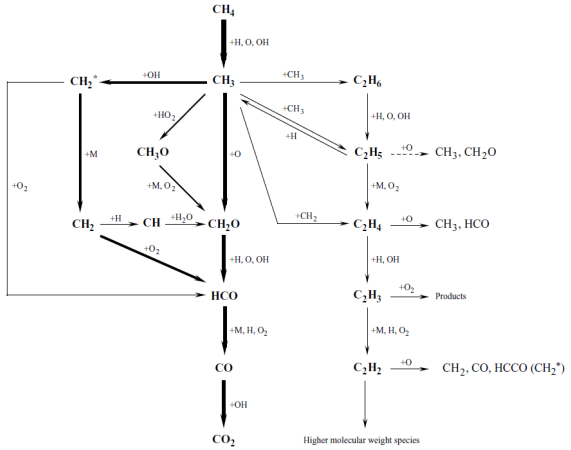
ABSTRACT

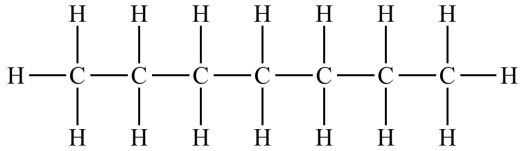
❖ Combustion Chemistry is complicate, where numerous elementary reactions and intermediate species are involved, especially for large fuels (gasoline, jet fuels …). People often simplify the chemistry by removing some reactions which have minor effect on the system to reduce the computation cost in simulations.

❖ In this work, we simplify the combustion chemistry using regularized regression models and feature selection methods, and also analyze the reaction network by different methods of clustering and community detection.

BACKGROUND AND APPROACH

Combustion Chemistry





- Complicate system with numerous reactions and species
  - ~10000 reactions and ~2000 species for practical fuels
  - Complex networks for chemical reaction fluxes
- Current Methods of chemistry mechanism reduction
  - Sensitivity analysis
  - Directed Relation Graph methods
  - Path Flux Analysis

Disadvantages: Require detailed information of concentration of all species during the reaction process

New Approach

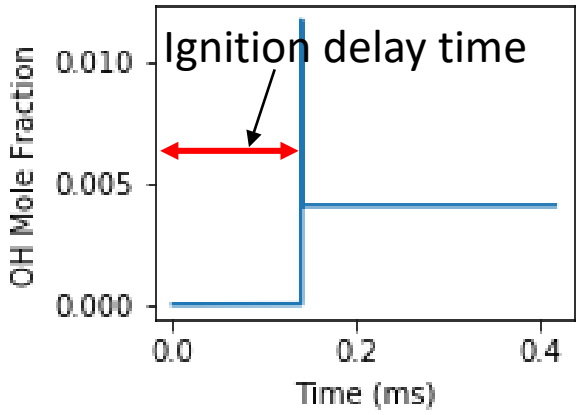
- Regression Models:
  - Linear regression models
  - LASSO regression
  - Ridge Regression
  - Forward stepwise regression
  - “Layer selection regression”
- Feature Selection:
  - From Ridge Regression
  - From LASSO
  - From Forward stepwise regression
  - From Layer selection regression
- Network Analysis:
  - Clustering
  - Label Propagation Algorithm

Regression Methods

Data:

- Butane reaction mechanism with **2457 reactions** and 230 species. We treat each reaction rate as a feature.
- Key outcome: **Ignition delay time** of mixtures with different initial conditions(Temperature, pressure, composition) by chemical reaction solver *Cantera* package in Python
- *Cantera* solves the huge ODE systems of the chemistry to simulate the chemical kinetics in experiments.

Cantera



- Reaction rate are calculated through *Cantera* for all 2457 reactions under different conditions as features.
- Train data: 2880 samples, test data: 600 samples

Task:

- Use $k_1, k_2, \ldots\ldots, k_{2457}$ to predict τ_{ig}

Regularized Regression:

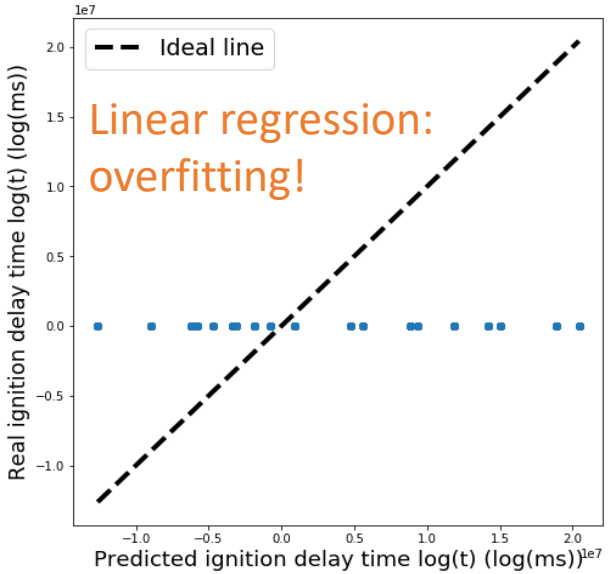
- Simple linear model has the overfitting issue.
- Use LASSO and Ridge to do shrinkage regression
- Use Forward Stepwise and Layer Selection to pick features

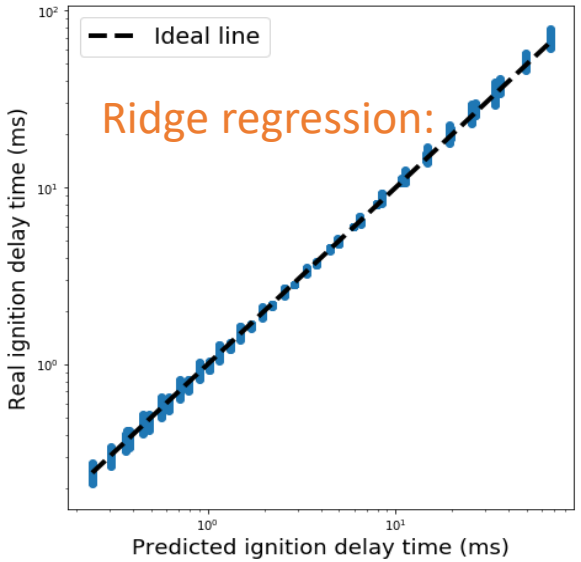
Feature Selection after Regression:

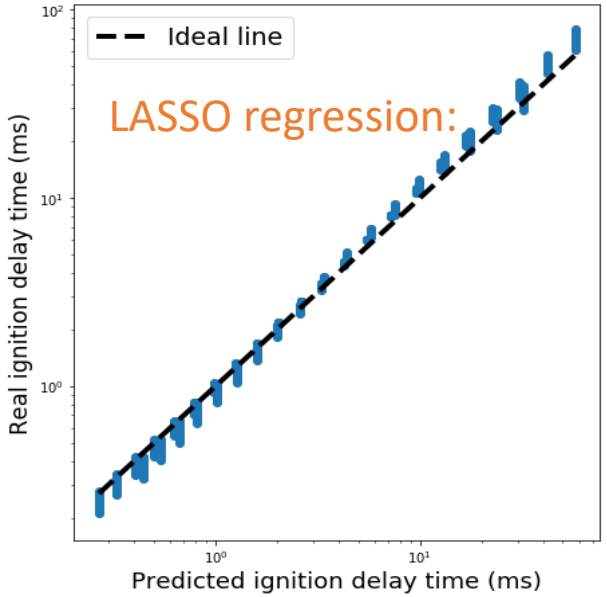
- Ridge Regression:
  - Select features with coefficients larger than a threshold
- LASSO Regression:
  - Remove features with 0 coefficients
- Forward Stepwise Regression:
  - Model selection based on BIC.
  - Pick features in the best model.
- Layer Selection Regression:
  - Categorize reactions into different layer with prior knowledge.
  - Run model selection based on BIC layer by layer
  - Optimally pick features in each layer to ensure ignition

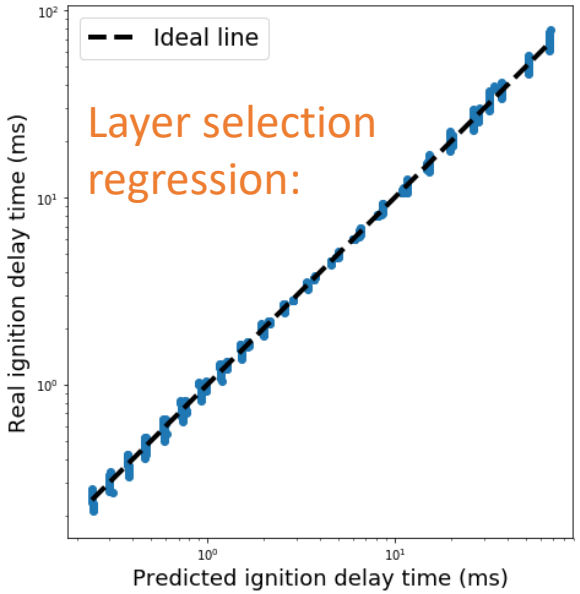
Regression Results

Regression







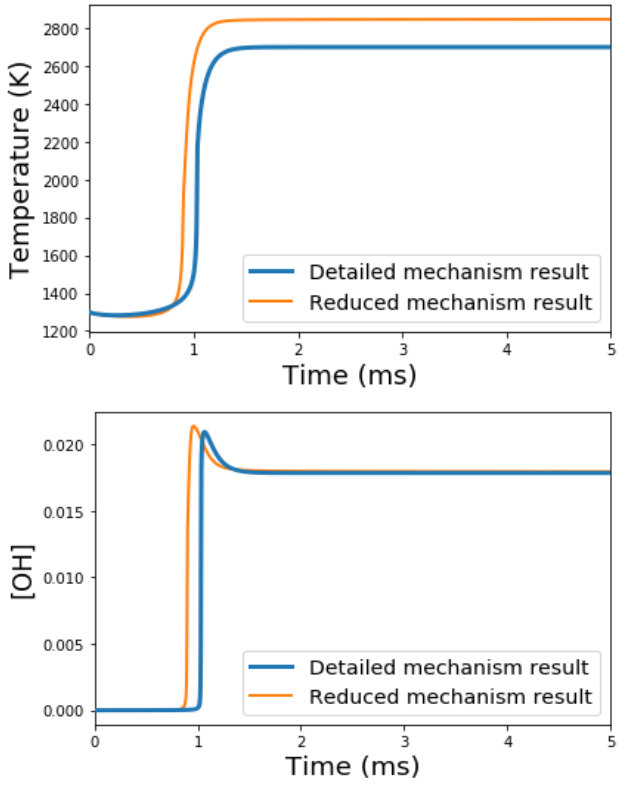


Feature Selection

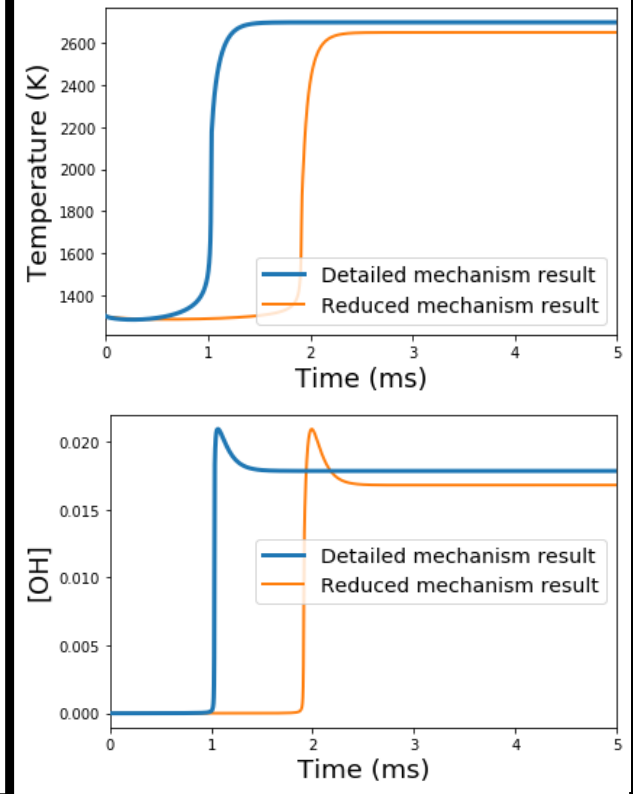
By comparing all regression models, we pick ridge regression and layer selection regression to select important reactions and reduce the mechanism.

Put the reduced mechanism back into *Cantera* and compare with the detailed mechanism:

Ridge: 1775 reactions, 225 species

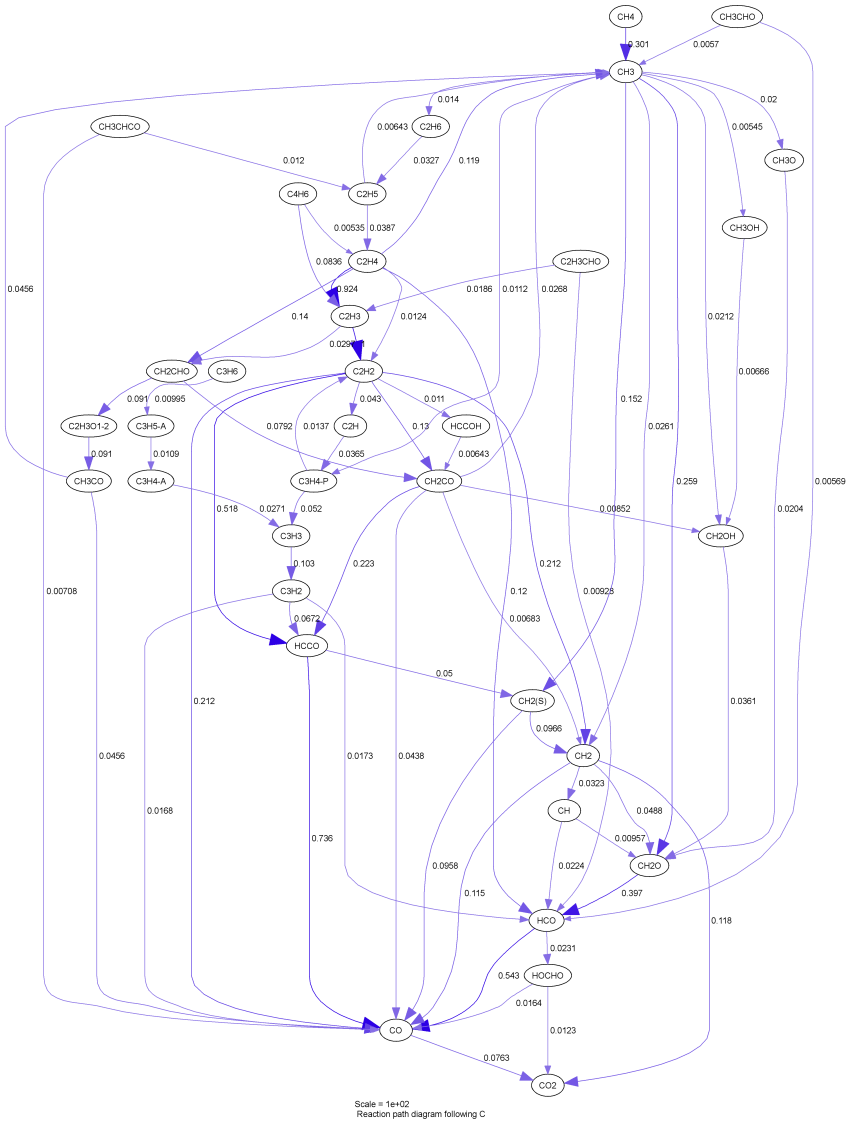


LSR: 591 reactions, 171 species

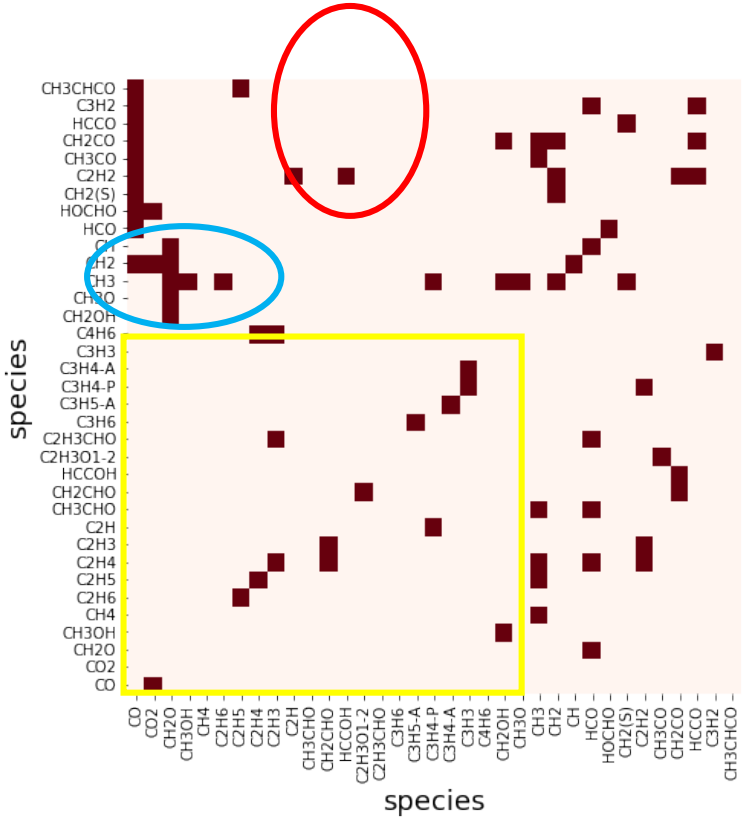


Network Analysis

Network of element flux in butane combustion



Clustering

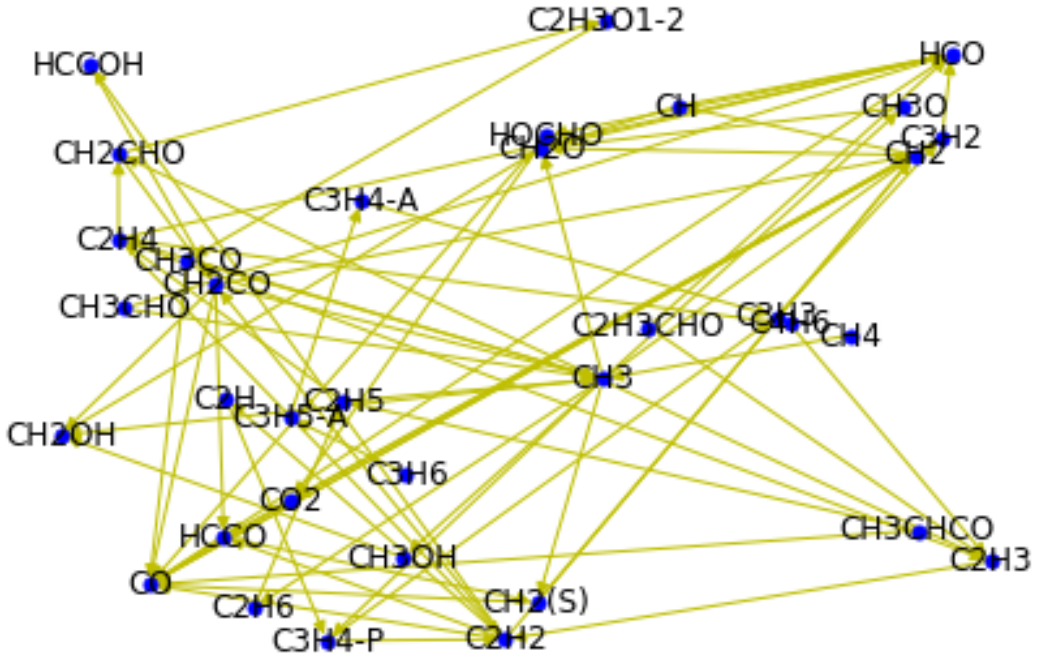


Three groups are divided:

- *HCO, HOCHO, CH_2, C_2H_2, CH_3CO, CH_2CO, HCCO, C_3H_2, CH_3CHCO* (all lead to CO production)
- CH_2OH, CH_3O, CH_3, CH_2, CH
- the rest species

Network Analysis

Community Detection



Label Propagation Approach

- Assign a random label to every nodes at first
- According to the labels of neighbors of each node, find its most probable label
- Iterate until converge