A Data-Driven Approach for Predicting Elastic Properties of Inorganic Materials
Camila Cendra (ccendra@stanford.edu)

Scope
- Materials discovery from first-principle computations is expensive[1,2].
- Identifying mechanical properties of new materials is crucial to determine their potential functionality.
- The elastic modulus measures a material’s resistance to deformation.
- We use machine learning (ML) methods to predict the elastic modulus (y) from common chemical properties, bypassing the need to use more expensive computational methods.

Data and Features
- Y data for training is elastic modulus
- 3039/537/632 train/dev/test split.
- Features (X) of the model are 135 descriptive attributes.
  ➢ 118 encode chemical composition
  ➢ 17 encode heuristic quantities
    • i.e.: electronegativity, valence electrons, atomic mass and size.
- X was standardized to zero mean and unit variance using training data.

Supervised Learning Models
1) Linear Regression (LR) + Regularization
   • Linear model that minimizes least squares loss while penalizing the size of coefficients (\(\omega\)).
   \[
   \min_{\omega} ||X\omega - y||^2 + \alpha ||\omega||^2
   \]
   Parameters
   • \(\alpha = 0.5\)

2) Multi Layer Perceptron (MLP)
   • Neural network with one hidden layer. Loss minimization with stochastic gradient descent.
   \[
   \text{Loss}(\hat{y}, y, W) = \frac{1}{2} ||\hat{y} - y||^2 + \frac{\alpha}{2} ||W||^2
   \]
   Parameters
   • \(\alpha = 0.0001\)
   • L_rate: 0.001
   • activation: relu
   • output: identity
   • 100 neurons in hidden layer

3) Random Forest Regressor (RFR)
   • Bootstrapped meta estimator that fits classifying decision trees.
   Parameters
   • Number trees: 100
   • Maximum depth: 15

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>train</th>
<th>test</th>
<th>10-fold CV</th>
<th>train</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>35</td>
<td>42</td>
<td>36 ±16</td>
<td>0.88</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>MLP</td>
<td>28</td>
<td>37</td>
<td>32 ±16</td>
<td>0.92</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>RFR</td>
<td>27</td>
<td>38</td>
<td>32 ±16</td>
<td>0.93</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

Feature Importance: as determined by 10-fold recursive feature elimination with cross-validation (RFECV), heuristic and compositional descriptive attributes are both useful.

Discussion
- Using descriptive attributes, which are readily obtained analytically for any given composition, we have predicted the elastic modulus of a diverse set of materials with high accuracy.
- 10-fold CV shows that the prediction performance of the ML models is consistent.
- Both heuristic and compositional features contribute to the models’ high performance.

Future Work and References
- Predict other crucial mechanical properties, such as shear modulus and fracture toughness, using the developed methodology.
- Implement advanced ensembling algorithms to achieve higher predictive accuracy.

References: