Machine Learning and Global Optimization for Materials Discovery

Logan Ward
Postdoctoral Scholar
Computation Institute, University of Chicago
3 October 2017
Machine Learning and Global Optimization for Materials Discovery

Logan Ward, Chris Wolverton
Graduate Student
Northwestern University
3 October 2017
Materials Engineering: A narrow introduction

Examples of “Aluminum”

Application

Key Properties
Conductivity Low cost Castable
Shapeable Strong ??

Alloy
Al + Mg Al + Mg,Mn Al + Si,Cu ???

How do we tailor materials for new technologies?

How can we do this quickly?
**Goal:** Accelerate design of materials

**Method:** Replace experiments with computers

**Many Established Tools:**
- Density Functional Theory
- Phase Field
- Finite Element Analysis
- Computational Thermodynamics

**Emerging Field:** Data-driven Models

\[ \sigma_Y = f(x) \]

Materials Data → Machine Learning → Predictive Model

FEA Image: http://www.icams.de/content/research/index.html
ML + Materials = “Materials Informatics”

Question: How can I do this, but for many more properties?
Materials Informatics Workflow

Collect > Process > Represent > Learn

\[ \Delta H_f = -1.0 \]
\[ \Delta H_f = -0.5 \]

\[ \Delta H_f = f(Z_A, Z_B) \]

<table>
<thead>
<tr>
<th>( Z_A )</th>
<th>( Z_B )</th>
<th>( \Delta H_f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>-1.0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>-0.5</td>
</tr>
</tbody>
</table>
Materials Informatics Challenges

**Collect**

**Process**

**Represent**

**Learn**

**Need:** High-quality data

**Question:** How to simplify using ML in MSE?

**Part 1:** Automated Crystal Structure Solution

**Part 2:** General-Purpose Representations for Materials

**Part 3:** Design of New Materials
Part 1: Automated Structure Solution

Acknowledgements: Kyle Michel
Why is this important?

One Reason: High-Throughput DFT

Automated calculation of properties

**Input:**
Structures of known materials

**Output:**
Large collection of materials data

DFT $\rightarrow V, \Delta H_f, ...$

DFT $\rightarrow V, \Delta H_f, ...$

DFT $\rightarrow V, \Delta H_f, ...$
**Why is this important?**

**One Reason:** High-Throughput DFT

Any HT search is limited by the availability of high-quality crystallographic data!

**Problem:** Structures unknown for many materials
Crystal Structure Databases

**Inorganic Crystal Structure Database**

Collection of the crystal structures of compounds

*Total entries:* 148,818

*Missing atoms:* 13,705 (~9%)

10³ – 10⁵ of materials not included in HT DFT Databases!

**Powder Diffraction File**

Database of Powder X-Ray Diffraction patterns

*Total entries:* 384,613

*Without structure:* 113,164

Proposed Solution: Solve them with FPASS
What is FPASS?

**Simple**: Single-step method to solve structure

**Accurate**: Uses both energy and diffraction pattern match

**Fast**: Constrains search with known symmetry

**Problem**: We had not used FPASS for *new* solutions

- Can we trust it in an automated scheme?
- Can we automate it?

FPASS Uses a Genetic Algorithm

Sum of two ranks:
- Pattern match
- DFT energy

By maintaining Wyckoff sites, we preserve symmetry
Solving Na$_x$Pb$_y$ Structures

Our starting point: Abstract from 1957

<table>
<thead>
<tr>
<th>Phase</th>
<th>Ideal formula</th>
<th>Space group</th>
<th>$a_0$ (Å)</th>
<th>$c_0$ (Å)</th>
<th>Pb atom positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Na$_5$Pb$_2$ (?)</td>
<td>R$ar{3}$m</td>
<td>5.54</td>
<td>23.15</td>
<td>6 (c): 0, 0, z; ...; $z = 0.07$</td>
</tr>
<tr>
<td>II</td>
<td>Na$_9$Pb$_4$</td>
<td>P6$_3$/mmc</td>
<td>5.47*</td>
<td>30.41</td>
<td>4 (e): 0, 0, z; ...; $z = 0.05$</td>
</tr>
<tr>
<td>III</td>
<td>Na$_{13}$Pb$_5$</td>
<td>P6$_3$/mmc</td>
<td>5.51</td>
<td>40.39</td>
<td>4 (f): 1/2, 3/2, z; ...; $z = 0.30$</td>
</tr>
</tbody>
</table>

Only missing feature: Na positions

Can we automate this?

Na positions with FPASS

New structure improve computed Na-Pb diagram

Computed T=0K Phase Diagram

Can we trust automated FPASS?

**Test Cases:** Strukturbericht
- Common structures
- ~100 compounds
- Variety symmetries, compositions, sizes

Can we solve them?

**Yes!**
New Solutions

And there are more on the way!

Ward, Michel, Wolverton, *in review*
Summary

**Goal:** Reduce the number of unsolved compounds

**Method:** First-Principles-Assisted Structure Solution

1. Created automated implementation
2. Validated against ~100 structures
3. Solved 13 crystal structures

**Result:** Improved accuracy of HT-databases

Ward, Michel, Wolverton. Physical Review Materials, *in review*
Part 2: General-Purpose ML Methods for Materials
How can one create problem-independent representations?
What is a representation?

*Set of quantitative attributes that describe a material*

\[
\text{Property} = f(\text{Attributes})
\]

Representation of material

Ex: \( Attributes = g(x_H, x_{He}, ...) \)

What does a representation need?

*Completeness:* Differentiate materials

*Efficiency:* Quick to compute

*Accuracy:* Capture important effects

*Diversity:* Many possible properties

How do we create “general-purpose” representations?
### Variety of Types of Materials Data

**How to differentiate materials?**

<table>
<thead>
<tr>
<th>Available Information</th>
<th>Element</th>
<th>Phase Diagram</th>
<th>Composition</th>
<th>Crystal Structure</th>
<th>μ-Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al, Zr</td>
<td></td>
<td># Eutectics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al₄Zr</td>
<td></td>
<td>Glass-Forming Ability</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example Properties</th>
<th>Dilute ( \Delta H_{mix} )</th>
<th># Eutectics</th>
<th>Glass-Forming Ability</th>
<th>( \Delta H_f )</th>
<th>( \sigma_Y )</th>
</tr>
</thead>
</table>

**Need:** A suite of general-purpose representations
Focus #1: Composition

*Property* = \( f(Composition) \)

<table>
<thead>
<tr>
<th>Property</th>
<th>Attributes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal Structure</td>
<td>( \text{VE, } \Delta X, \ n_{av}, \Delta n_{ws}^{1/3} )</td>
<td>Kong et al., 2012</td>
</tr>
<tr>
<td>Band Gap</td>
<td>( \Delta X, Z, T_m, R, n_{av} )</td>
<td>Srinivasan &amp; Rajan, 2013</td>
</tr>
<tr>
<td>Formation Energy</td>
<td>( \Delta X, Z, n_{s</td>
<td>p</td>
</tr>
<tr>
<td>Melting Point</td>
<td>( Z, m, n, r_{\text{cov}}, l, X, \ldots )</td>
<td>Seko et al., 2014</td>
</tr>
<tr>
<td>( \Delta H_f: \text{Rocksalt} - \text{Wurtzsite} )</td>
<td>( \text{IP, EA, } r_s, r_p, \ldots )</td>
<td>Ghiringhelli et al., 2015</td>
</tr>
</tbody>
</table>

**Observations:**
- Different properties, different attributes
- All based on elemental property statistics

**Our Strategy:** Create set that includes all of these and more
General-Use Attributes

**Elemental Property Stats.**: Mean $T_m$, Range $Z$, ...

*6 Statistics*: Mean, variance, max, min, range, mode

*22 Elemental Properties*: $Z$, $EN$, Row, Column, Radius, ...

**Stoichiometric**: # Components, $\|x_Z\|_p$

**Electronic Structure Based**: Fraction $p$ Electrons, ...

**Ionicity**: Can form Ionic, % Ionic Character, ...

Simple Example: Is it a Metal?

**Task:** Given composition, $E_g > 0$?

**Training Set Dataset:** 3000 entries from the OQMD

**Simple ML Model:** Accuracy 84.1%

Game: palestrina.northwestern.edu/metal-detection/
Application to the OQMD

Dataset: 240000 DFT Calculations (OQMD.org)

$\Delta H_f$

Ref: Ward et al. npj Comp. Mat., (2016) 28
Predicting Glass Forming Ability

Application: Metallic Glasses

Goal: Predict glass-forming ability

Dataset: Landolt-Börnstein
- 6836 experimental measurements
- 295 ternary systems
- Binary property: [Can Form Glass] | [Cannot Form]

Model: Random Forest
- 90% accurate in 10-fold cross-validation

Ref: Ward et al. npj Comp. Mat., (2016) 28
Predicting Glass-Forming Ability

**Test:** Remove Al-Ni-Zr data from training data, try to predict

**Measured**

**Predicted**

Same representation, very different material.

Ref: Ward et al. npj Comp. Mat., (2016) 28
Focus #2: Crystal Structure

Our Approach:
Voronoi-tessellation-based attributes

Atomic Characteristics:
1. Element identity
2. Coordination number
3. Bond length
4. Cell size
...

Atomic Characteristics + Descriptive Statistics = 275 Attributes

Ref: Ward et al. PRB. (2017) 024104
Learning Rate Comparison

**Dataset:** 32k DFT $\Delta H_f$ from the OQMD

**Test:** Remove 1000, train on $N$ remaining

Cross-validation is great, but does not model real use

Common Method: Prototype Search

1. Select a crystal structure
2. Evaluate all possibilities with DFT
3. Select only stable ones

Challenge: Computational cost

Possible Solution: Guide with ML

Ref: Ward et al. PRB. (2017) 024104
Ranking Candidate Materials

**Training Set:** 32k entries from OQMD

**Test:** Select top 100 entries from test set

Ranks entries better than existing methods, good choice for accelerating combinatorial searches
Future Steps: Open Benchmarks

**Challenge:** Which method is best for a certain dataset?

**Our proposal:** Develop a set of benchmark challenges

**Approach:** Host data, models using MDF and WholeTale

**Current Progress:** Compared several composition-based models

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full OQMD</td>
<td>0.161</td>
</tr>
<tr>
<td>Deml 2016</td>
<td>0.170</td>
</tr>
<tr>
<td>Meredig 2014</td>
<td>0.072</td>
</tr>
</tbody>
</table>

WholeTale.org
Summary

Collect | Process | Represent | Learn

Composition-Based Attributes | Crystal Structure Attributes

How do I use this to find a new material?
Part 3: Using ML To Find Materials

Acknowledgments:

Liquidmetal: S. O’Keeffe, J. Stevik, G. Jelbert

NU: M. Aykol, K. Kim, J. He
Example Design Problems

Designing Metallic Glasses Alloys

1. Optimizing commercial alloys
2. Locating new glasses with sputtering

Discovering Quaternary Heuslers
**Application: Bulk Metallic Glasses**

*Metals with Amorphous Structures*

- No Dislocations -> High $\sigma_Y$
- Glass Transition -> Net Shape Casting
- No Grain Boundaries -> Low Hysteresis

**Applications:** Surgical tools, flight control surfaces, ...

**What are the tradeoffs?**

Source: liquidmetal.com

Source: Wikipedia.org
Problems with BMGs

Main Problem: Metastability of glassy phase

Few Alloy Systems Known to Form Glasses
<2% possible ternaries in Landolt-Börnstein handbook

Small Critical Casting Thicknesses
Typical thickness < 10 mm

Small Time/Temperature Processing Window
BMGs may crystallize during processing

Design Issue: Lack of Composition -> Properties Links

Our Proposed Solution: Data analytics

Goal: Optimize $D_{\text{max}}$ and $\Delta T_x$
Materials Informatics Workflow

Collect  Process  Represent  Learn

\[ \tilde{X} = f(\text{Composition}) \]

“General Purpose” Attributes
+ Cluster Packing Models
+ DFT \( \Delta H_f \) from OQMD

- GFA: 6315 entries
- \( D_{max} \): 5916 entries
- \( \Delta T_x \): 621 entries

- GFA: Random Forest
- \( D_{max} \): Random Forest + Additive Regression
Task #1: Tuning Known Alloys

**LM105 Optimization**

- **LM105-Opt1**: $\text{Zr}_{55.0}\text{Ti}_{2.0}\text{Cu}_{23.0}\text{Ni}_{7.5}\text{Al}_{12.5}$
- **LM105-Opt2**: $\text{Zr}_{47.0}\text{Ti}_{3.0}\text{Cu}_{23.5}\text{Ni}_{11.5}\text{Al}_{15.0}$
- **LM105-Opt3**: $\text{Zr}_{47.0}\text{Ti}_{2.0}\text{Cu}_{23.0}\text{Ni}_{18.0}\text{Al}_{10.0}$
- **LM105-Opt4**: $\text{Zr}_{60.0}\text{Ti}_{5.0}\text{Cu}_{17.5}\text{Ni}_{10.0}\text{Al}_{7.5}$
- **LM105**: $\text{Zr}_{52.5}\text{Ti}_{5.0}\text{Cu}_{17.9}\text{Ni}_{14.6}\text{Al}_{10.0}$
Tuned Alloys: Pareto Analysis

Predicted alloys have superior properties

Results: Improved design space
**Task #2: Discovering New Alloys**

**Search Space:** Ternary alloys of 53 elements

- 26 million ternary alloys
- 945k alloys: $L_1$ distance > 30at%
- 290k alloys: $P(Glass) > 95$
- 57k alloys: $\Delta T_x > 66$ K
- 38k alloys: $D_{max} > 1$ mm

**Run Time:** ~ 2 days
Chosen System: Cu-Hf-[Mg,Ti]

$D_{max}$

$\Delta T_x$

Cu-Hf-Ti

Cu-Hf-Mg
Cu-Hf-[Ti/Mg]: Current results

- Sample crystallized and/or shattered during injection molding
  - $\text{Cu}_{60}\text{Hf}_{25}\text{Ti}_{15}$ could be formed with copper mold casting

Mg ignited during arc melting

Lessons Learned:
- Integrate data from target process
- Screen based on processing technique

**Example #2: Finding New Glasses**

**Goal:** Find new sputtering glasses

**Step 1:** Include processing information

**Step 2:** Scan 2.4M alloys (~5 min) -> Selected Co-V-Zr
Decent agreement. ML model “Zr-lean”, but close enough for success
Repeat, with Improved Model

Initial Model
Add Co-V-Zr
Improved Accuracy
Goal: Find more Quaternary Heuslers (QHs)

Why quaternary Heuslers?

Ternary Heuslers have great properties, so a 4th degree of freedom could be better?

Problem: ~3M possible combinations

Solution: Guide search with ML

Ref: Kim et al., in preparation
Materials Informatics Workflow

- Collect
- Process
- Represent
- Learn

**Input:** Crystal structure
**Output:** Stability

**Training Data:** OQMD
**Entries:** 410981
96160 QHs

**Δ 𝐻ₚ**

**Voronoi-Tessellation Method**
275 attributes

Random Forest
Step 1: Select Training Set, Validate

Training data: All data available, or only similar materials?

Small amount of QHs available: Select as much data as possible

More QH data: All equivalent

Bottom Line: Training on all available data is advantageous
Step 2: Find New Heuslers

ML success rate:
55 / 909 (6%)

Original search:
353 / 96189 (0.3%)
Automated Crystal Structure Solution
 ✓ Validated FPASS method, created automatic framework
 ✓ Solved 13 crystal structures

General-Purpose Representations
 Using Composition: DFT $\Delta H_f, E_g, V$; Glass-forming ability
 Using Crystal Structure: 2x better than existing approaches

Example applications:
 ✓ Improved $\Delta T_x$ of 2 commercial BMGs alloys
 ✓ Discovered many new metallic glasses
 ✓ Found dozens of new quaternary Heuslers

Open Source Software: Magpie, Matminer
 http://bitbucket.org/wolverton/magpie
 http://github.com/hackingmaterials/matminer
Acknowledgements

Funding:
• CHiMaD/NIST
• NDSEG Fellowship Program
• Weertman Fellowship, Ryan Fellowship

Wolverton Group
Vinay Hedge, Murat Aykol, Max Amsler, Scott Kirklin, Yang Yu, Kyoungdoc Kim

Choudhary Group
Ankit Agrawal, Rosanne Liu, Amar Krishna

Liquidmetal Technologies
Stephanie O’Keeffe, Joseph Stevik, Glenton Jelbert

NIST/USC/SLAC
Apurva Mehta, Fang Ren, Jason Hattrick-Simpers, Travis Williams