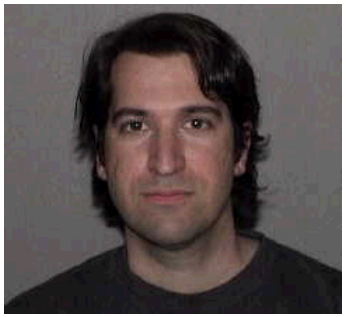
SAND2017-8375C

Thermoelectric Materials Discovery: A Domain for ML/DL Search?

2017 Sandia MLDL Conference

Kurt Larson, 6362



Peter Sharma, 1867

Phillip Lewis, 6362

Motivations

- Designing materials with a complicated combination of properties is the most fundamental problem in materials science and engineering
  - Extremely difficult to achieve in practice
  - Most functional materials discoveries are accidental and explained after the fact by modeling
- Seek a way to accelerate materials discovery

Why Thermoelectrics?

- Thermoelectric materials directly convert heat into electrical power with no moving parts
- Several DOE mission areas use them
- It is extremely difficult to find materials with the right combination of transport properties that lead to high thermoelectric conversion efficiency.
- New discoveries
  - Are few and infrequent
  - Seldom occur outside of semiconductors
  - Could significantly improve mission applications
- A reasonable exemplar for other material science applications

Search Space / Data Support

- Good thermoelectrics are characterized by multiple crystal transport properties.
- Databases exist of 500K+ known materials with many properties, not including thermoelectric properties
- <50K known materials with measured thermoelectric properties
- Formulating a new material for testing is exceedingly expensive
- A first-principles calculation to predict properties exists, but is too expensive for broad search
- Several less-expensive and less-accurate models exist, which can be run for candidate materials

Prior Work

- Regression and Random Forest models for thermoelectric property prediction, mostly semiconductors – recommendation engine interfaces
  - Sparks et al, 2016; Gaultois et al., 2016
- ML for materials science at Sandia
  - Truman Fellow Julia Ling (departed to industry)
- Rapidly increasing research in materials informatics
- Bioinformatics analogues

Elsevier charges $29.90 for this figure

Sparks et al., 2016, Acta Materialia, Fig. 3

- Our proposed work differs:
  - extend to material classes outside of semiconductors
  - Make use of ML methods not yet tried

Algorithm Idea

- Sequential Learning Neural Network Paradigm
  - Train NN on existing materials & thermoelectrics data
  - Predict where good thermoelectrics might be
  - Calculate properties of predicted materials with some combination of cheap and expensive methods
  - Re-train NN, adding in new calculated values
  - Iterate until NN is reasonably predicting values the model subsequently calculates
  - Submitted as Idea to FY18 LDRD, not selected.

- Are there better approaches to this prediction problem?
- Is this problem a direct analogue for something that someone else at Sandia is already doing?
- Any suggestions on where to go for funding?