Stefan Knapik

MDS Project Proposal: Predict properties of aluminum alloys

Objective

Build an app that can predict the mechanical properties of an aluminum alloy given its composition (or an app that can determine potential alloy compositions based on desired mechanical properties).

1. Multimodal data generation and collection

Material test data is available online. The two main datasets I found come from Materials Cloud Archive and Taylor & Francis Online. The associated works are listed at the bottom of this page.

2. Featuring engineering

Aluminum alloy composition is characterized by proportions of various elements including aluminum, iron, manganese, silicon, magnesium, titanium, copper, chromium, zinc, vanadium, and zirconium. The mechanical properties of aluminum alloys include yield strength, tensile strength, and elongation at break. Features not captured in the above datasets that could potentially be incorporated from other datasets are heat treatment, process parameters, corrosion resistance, and fatigue strength.

3. Dimension reduction

Dimensions could be reduced by only considering certain alloying elements or by segregating alloying elements into families if their concentrations are strongly correlated.

4. Reduced order modeling / Regression and classification

Singular value decomposition, principal component analysis, or proper generalized decomposition may be used to reduce the order of the model. Chapter 5 will inform this portion of the project when the time comes. Supervised machine learning models can be trained, validated, and tested. The model that performs the best will be used.

6. System and design

Aluminum alloy development may be accelerated if the underlying patterns can be captured using MDS techniques. The performance of new alloy compositions could be predicted which would save time and money on real-life material testing. Aluminum alloy development is important because aluminum alloys are distinguished by their light weight which makes them suitable to energy saving aerospace and automotive applications.

References

Olivia P. Pfeiffer, Haihao Liu, Luca Montanelli, Marat I. Latypov, Fatih G. Sen, Vishwanath Hegadekatte, Elsa A. Olivetti, Eric R. Homer, Aluminum alloy compositions and properties extracted from a corpus of scientific manuscripts and US patents, Materials Cloud Archive 2021.80 (2021), doi: 10.24435/materialscloud:vx-fy.

Tamura, Ryo, et al. “Materials Informatics Approach to Understand Aluminum Alloys.” Science and Technology of Advanced Materials, vol. 21, no. 1, 2020, pp. 540–551., doi:10.1080/14686996.2020.1791676.

Portfolio optimization by Sharpe ratio maximization (project continuation)

From Optimization to Machine Learning

The project in its current form is simply an optimization over a singular time period. We get historical data for a certain number of days, and we find the portfolio that would have performed the best during those days.

There is no reason to expect the model to generalize to the data we do not have yet. The most important step forward in my opinion is that we develop code for performing gradient ascent using stochastically selected lookback windows from a much larger set of historical data.

For example, we can get financial information spanning the past 10 years. We can choose a batch size of 40 and a lookback window of 50 days. This would mean that when performing gradient ascent, instead of backpropagating from one 50-day time period, we will be using 40 50-day time periods. It also means that with each iteration of gradient ascent, we will be choosing a new random sample of 40 50-day time periods out of the total 10 years' worth of collected data. Of course, some data will be reserved for validation and testing, but this strategy should produce a much more robust model. This methodology would be considered machine learning.

Objective Function Alternatives

We can perform gradient ascent on any differentiable function!

- Modified Sharpe ratios
  - Recent information weighted heavier
  - Lagging Sharpe ratio relative to corresponding input financial data
- Sortino ratio - negative deviation is the proxy for risk instead of standard deviation
- Portfolio diversification - not sure how to quantify this but I know that people do it

Variation of Hidden Network Architecture

- Deep model with multiple neural layers
- Recurrent networks
  - Simple
  - Long Short-Term Memory
  - Gated

Backtesting against other traditional trading strategies

We can compare portfolio performance in terms of Sharpe ratio and other common indicators.