Computational Alloy Design and Discovery using Machine Learning

Yadu Krishna Choyi

Comp4560

Supervisor: Nick Birbilis

Motivation

“Worldwide there are millions of materials available commercially that are characterised by hundreds of different properties. Using traditional techniques to explore the information we know about these materials, to come up with new substances, substrates and systems, is a painstaking process that can take months if not years,” Gareth Conduit, the Chief Technology Officer at Intellegens

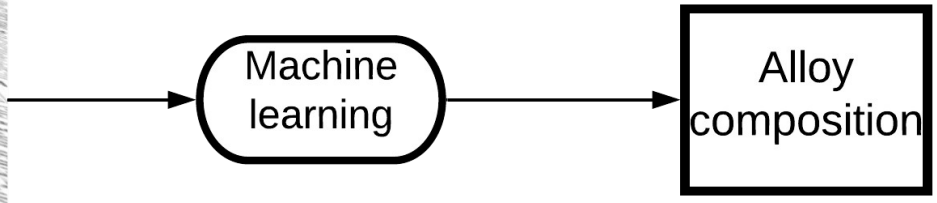
What is alloy designing and why do we need it?

- Knowledge-guided approach
- In the field of engineering we are typically using only up to 1000 different types of metallic alloys on a regular basis.
- This is a very tiny fraction of potentially interesting alloys

Current Methodology

- Semi-empirical chemical rules.
- Other approaches are to reflect which type of property we actually aim to design.
- Using already existing data to learn machine learning models that can accurately predict alloy compositions.

Objective



Objective

- Generate aluminium alloy composition using machine learning.
- Aim to generate aluminium alloy composition such that certain criteria are met.

Criteria

- Percentage of Mg ≈ 3.5% - 6.0%, Mn ≈ 0.1% - 0.6%, Fe ≈ 0.1% - 0.4%, Cr ≈ 0.0% - 0.3%, Ti ≈ 0.0 % - 0.3%, Sr ≈ 0.0% - 0.5%,Zn ≈ 0.05% - 0.75%,Zr ≈ 0.0% - 0.5%,Cu ≈ 0.0% - 0.7%, Ag ≈ 0.0% - 0.2%, Si ≈ 0.0% - 0.5%, Ni ≈ 0.0% - 0.2%, Ca ≈ 0.0% - 0.2%, Ge ≈ 0.0% - 0.4%, Nd ≈ 0.0% - 0.4%, Ce ≈ 0.0% - 0.4%
- Aim to minimize DoS value

Degree of sensitization(DoS)

- Sensitization in 5XXX aluminum alloys is a problem characterized by the gradual formation and growth of beta phase (Mg2Al3) at grain boundaries, which increases the susceptibility of alloys to inter granular corrosion (IGC) and intergranular stress-corrosion cracking (IGSCC).
- 5XXX series aluminium is commonly used in hull construction for ships.
- The DoS value gives a measurement of the extent of sensitization in such alloys.
- High DoS value implies less corrosion resistance, and vice-versa.

Methodology Overview

Data

Extracted from ‘A Survey of Sensitization in 5xxx Series Aluminum Alloys’ by R. Zhang,* S.P. Knight,* R.L. Holtz,** R. Goswami,** C.H.J. Davies,*** and N. Birbilis‡.

- Consists of 458 records

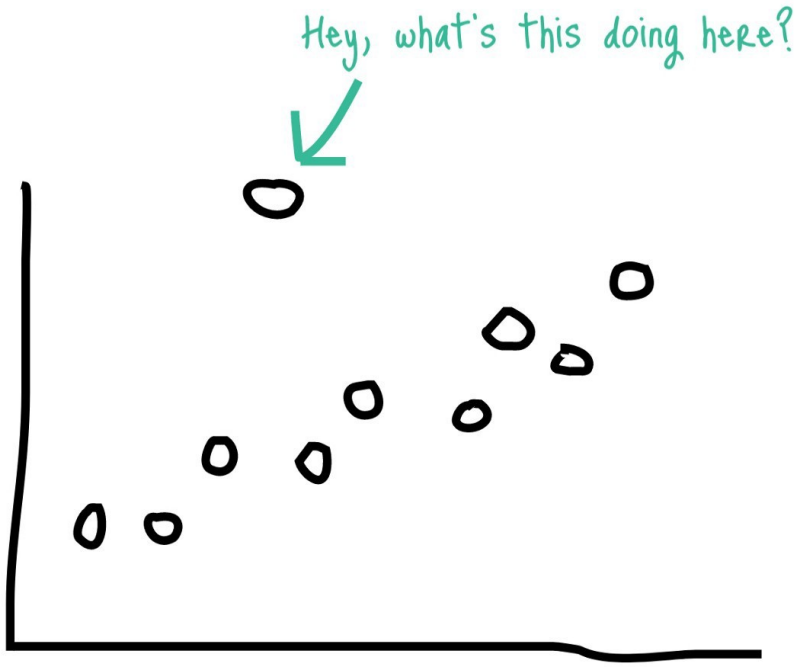
Data

- Number of data points = 458
- Number of input features = 21
- Number of output features = 1

A	B	C	D	E	F	G	H	I	J
Alloy Name	Sensitisation Time (days)	Sensitisation Temperature (Celcius)	Recrystallised (0 = no, 1 =yes)	Temper	Al	Mg		Ce	DoS
Al-4Mg/150	7	150	0	1	95.874	4	...	0.00	24.92
Al-4Mg	7	25	0	1	95.874	4	...	0.00	5.24
Al-4Mg-0.1A	7	150	0	1	95.874	4	...	0.00	30.13
Al-4Mg-0.1A	7	25	0	1	95.774	4	...	0.00	5.59
Al-4Mg-0.1S	7	150	0	1	95.874	4	...	0.00	10.75
Al-4Mg-0.1S	7	25	0	1	95.774	4	...	0.00	5.56
Al-4Mg-0.1T	7	150	0	1	95.784	4	...	0.00	18.87
Al-4Mg-0.1T	7	25	0	1	95.784	4	...	0.00	6.27
Al-4Mg-0.1Z	7	150	0	1	95.779	4	...	0.00	34.70

Data pre-processing / removing outliers

- Machine learning algorithms suffer in terms of their performance when outliers are not taken care of
- How to get rid of outliers?
- Solution = Isolation forest



Isolation Forest

- Isolation Forest is built on the basis of decision trees.
- Identifies anomalies instead of profiling normal data points.
- Anomalies are more susceptible to isolation and hence have short path lengths.

Approach

- Step 1 : Select a machine learning model that can predict the value of DoS.
- Step 2 : Tune the model.
- Step 3 : Generate alloy composition data to be fed to the model based on criteria.
- Step 4 : Feed generated data to the model and obtain DoS value to evaluate the alloy composition.
- Step 5 : Extract optimum alloy composition based on the value of DoS.

Model creation technique

- Supervised Machine Learning
- Neural Network

Which technique?

- Is supervised machine learning more suitable than a neural network model for the task in hand?
- Answer : Yes

Why supervised machine learning?

- Small size of dataset
- Faster execution time
- Higher accuracy

Machine learning model vs Neural network model

- Machine learning
- Model = xgboost
- Average r2 score = 0.94
- Average training time = 0.58 seconds

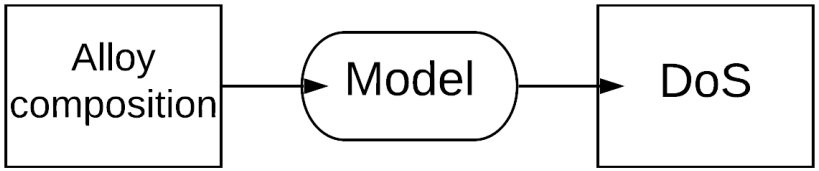
- Neural network
- Model = Simple feed forward neural network with 2 hidden layers
- Average r2 score = 0.52
- Average training time = 12.98 seconds

Xgboost vs other supervised learning algorithms

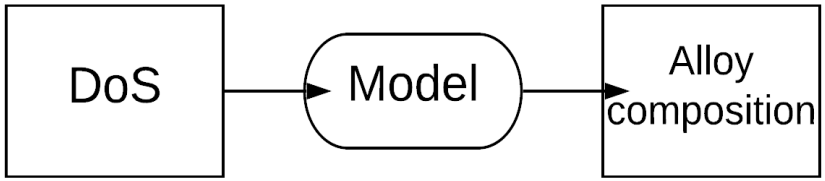
Linear	Nearest Neighbor	Decision Tree	SVR	Xgboost
R2 = 0.125	R2 = 0.657	R2 = 0.707	R2 = 0.820	R2 = 0.943

Recap

- What we have
- A good supervised machine learning model

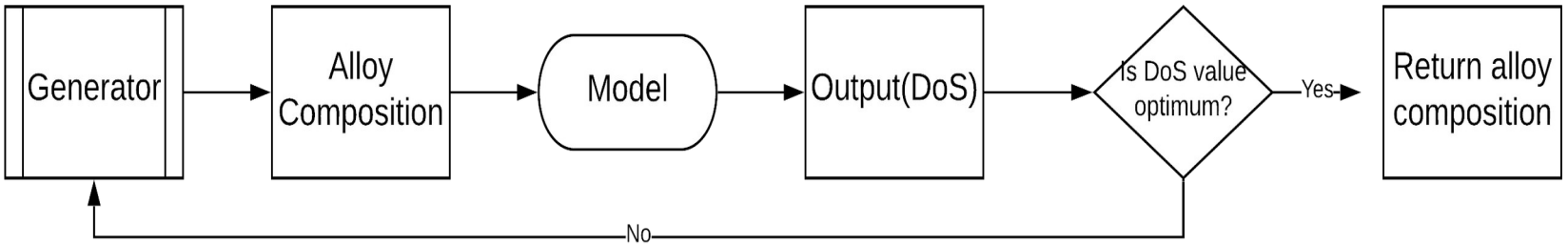


- What we want
- Alloy composition



Technique to generate composition

- Main idea is screening.
- Generate random composition satisfying all criteria.
- Use trained model to screen all generated composition and find best result.



How to generate random data?

- Two methods

- 1: Randomly select a the composition with range equal to criteria.

- 2: Randomly sample a point from the same feature space as that of the training data.

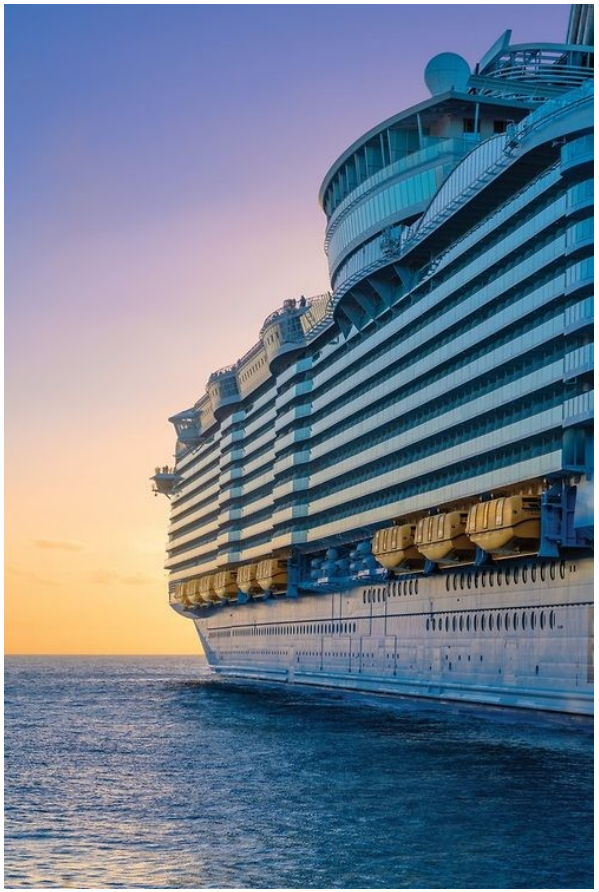
Result

- An average of around 10^5 random compositions are screened to select the one with the most optimum DoS value.
- DoS value of resulting composition is in the range (0.1 – 1.1).
- Comparing with average DoS value of original dataset which is 26.9, the generated composition seems to outperform existing alloy composition by a large margin.
- Example composition generated: Dos = 1.04

  Al=94.7, Mg=4.01, Mn=0.49, Fe=0.35, Cr=0.11, Si=0.2

Conclusion

- Marine engineers and naval architects recognise 5XXX series aluminium as an advantageous material in shipbuilding and the fabrication of components in offshore platforms. The lightweight, superior mechanical properties, and corrosion resistance of aluminium alloys has dictated their use in many of these applications.
- Using existing data to generate new alloy composition using machine learning can drastically decrease designing time.
- These new alloys can also find use in other fields such as the automotive industry.



Thank You!