Conclusions

This paper delivers my study of application of machine learning to improve the process of alloy discovery. It focusses not only on building and evaluating machine learning models for predicting properties of copper alloys using their composition but also provides a solution to creating a system which can be used to efficiently generate new copper alloys . The structure that has been followed in this paper is designed with the intention that it can be applied in the discovery of alloys of metals other than copper, starting from data pre-processing followed by model selection and ultimately creating an inverse model that can generate candidate alloy compositions with desired physical properties.

This use of machine learning as a tool for building predictive models as opposed to deep learning models (as done in (1)), has proven to be an effective way of identifying which features of alloys have a large contribution to their tensile strength and thermal conductivity as well as ones which don’t (refer section). This is a significant strength of this paper which can guide the feature engineering process in future studies on related topics. This unambiguous nature of machine learning models, as opposed to the black box like nature of deep learning models, bring significance the findings of this paper. Given the demonstration of powerful predictive abilities of neural network like models, we speculate the possibility of a combination of the two for this task. We strongly believe that the predictive abilities of deep learning architectures can be combined with the insights of the relevant features identified in this paper. We encourage that further research be devoted to this domain.

We believe that the fact that we were able to generate significantly accurate predictions for most of the alloys with no indications of overfitting demonstrates the existence of a relationship between alloy compositions and their physical properties. This knowledge can help the development of more sophisticated systems for alloy discovery using the power of artificial intelligence to drive the process.

Quick prediction time.

Limitations

Although, there is evidence that our alloy generation system provides good potential candidates, there is still a lot of scope for improvement.

The performance of the current models, although impressive, are not sufficient to be deployed in a real-world system, especially for thermal conductivity.

The dataset used to train the models in this paper contains alloys containing only 26 different elements (excluding copper). And the proposed system only generates combinations of existing alloys. So the generated alloys only come from this small subset of the potential search space. For example, Alloys containing carbon s (ref) have shown to have many of the desirable properties of good alloys but are not included in the dataset, therefore, copper alloys containing Carbon cannot be generated by our system.

Future Work

The insights of important features acquired from the Random Forest models can be used to engineer new and powerful indicators of the tensile strength and thermal conductivity and be applied to more sophisticated models for improving the performance of the current model.

We believe that the thermal conductivity model’s performance could be significantly improved by the addition of more data and is essential before the system can truly provide utility in the space of alloy discovery.

Currently, only a few of the alloy compositions generated by the inverse model have been inspected. Further analysis of the quality of the generated alloy by the inverse modelling technique is required before the system can be deployed as a solution for copper alloy generations, either by investigating existing copper alloy databases or by a process of developing the candidate alloys in a controlled environment (such as a lab) and evaluating their tensile strength and thermal conductivity properties.