Video presentation

Hi, I’m Raymond Wang.

Hi, I’m Dawning.

This is the video presentation of our final project for Data\_Science 423.

Page1

Our project is about understanding the composition-property relationships in alloy steels using machine learning.

Page2 Motivation

Alloys are a family of versatile materials widely used in modern industry. Even slight changes in the alloy composition could lead to a dramatic change in its physical properties. However, the state-of-the-art computational methods are typically computationally expensive to model the composition-property relationships. It is therefore our aim to use machine learning to address this challenge.

Page3 Overview

Hereby we present a reliable and efficient way to directly predict alloy steel properties from their elemental composition using machine learning.

After initial data acquisition and preprocessing, we analyzed our data and benchmarked different machine learning model performance. Then we used the best model to further study the composition-property relationship from our dataset.

Page4 Data acquisition and preprocessing

We scraped 855 instances of alloy steel data from an unstructured online resource, MatWeb. After converting the data type from string to floating point numbers, we finally obtained 855 instances with 13 physical properties, and 10 features which are the % weight of each element.

Page5 Data analysis

Now let’s take a look at our data. We plotted the correlation heatmap of these properties. The more red a block is, the stronger positive correlation these two properties have, and vice versa. We do see some interesting data structures here. For instance, layered structure here and dispersive data distribution here. However, we could only get qualitative relationship from here while we want more quantitative understanding.

Page6 Machine learning model

Then we built our machine learning models to predict different material properties using both elemental composition and other properties as input. We benchmarked 10 different algorithms as listed here, we used grid search for hyper-parameter tuning and 5-fold cross-validation. The results are show on the right. Circles and squares represent training and testing errors, respectively. In general, we found that XGBoost outperforms the other algorithms.

Page7 Learning with XGBoost

Therefore, we chose XGBoost for a more challenging task, where we only use elemental composition as input and try to predict alloy steel properties. The R2 scores of our training and testing set are shown here. Most of them actually did not show promising results, where both over-fitting and under-fitting situations are observed. However, XGBoost did a good job in predicting thermal conductivity. In order to validate our findings, we plot the learning curve here. As the number of training size increases, cross-validation score also increases. Therefore, having more training data would make our model even better.

Page8 Conclusions

In conclusion, we found that machine learning could provide quantitative insights of alloy steel composition-property relationships. XGBoost outperformed the other machine learning algorithms in our benchmark, and was particularly successful in predicting thermal conductivity. Additionally, the model performance could be further improved by having more training data.

End

This is all about our project, thank you so much!