Machine Learning Engineer Nanodegree

Capstone Project

Peerapon Wechsuwanmanee
December 21st, 2018

Definition

Project Overview

To characterize material properties for further analysis, e.g., Finite-Element Method (FEM), one shall distinguish the effect of each environment variables from the experimental results. There are analytical approaches to decouple them but it still needs expert's decision during the process. This project aims to develop a machine learning model to predict material's mechanical behavior which decouple the effect of temperature and speed, i.e., strain rate.

Problem Statement

In order to evaluate material's strength, one should characterize its stress-strain curve (flow curve) via uniaxial tensile test. For those who do not familiar with stress σ and strain ε, they are force and displacement respectively which are normalized by geometrical values. These values are normally obtained by the test at room temperature (RT; 293 K) and quasi-static loading strain rate (QS; 2.78×10^{-4} s^{-1}). The test is performed according to ASTM standard. The sample geometry can be found in Fig. 1. During the test, the sample is fixed on either left or right end then pulled from another side. Its force and displacement in process is recorded until the sample fractures. These information is converted to stress and strain before exporting as the final output.

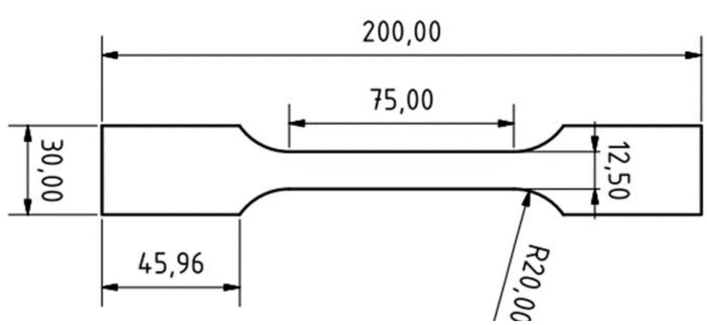

Figure 1: Geometry and dimension of a tensile test sample

However, it should be noted that in real-life application, the material shall operate in various environments, i.e., different temperatures T and/or different strain rates $\dot{\varepsilon}$. The material responses in these environments also differ from its behavior in standard environment.
To identify the effects, high speed uniaxial tensile tests and elevated temperature tensile tests are taken into account. The elevated temperature tensile tests are performed in QS loading speed. Thus, it takes no effect from the strain rate since the loading speed is the control factor in the machine during the process. However, in the high speed tensile test, the temperature in the sample cannot be controlled appropriately. Due to the fact that there is always heat generated inside the material as the loading is applied, the very high speed ones, i.e., at 1 x 10^{-1} and 1 x 10^{-2} s^{-1}, do not allow adequate time for heat dissipation to the material's ambient. Hence, the high speed tensile test results include coupled effects between temperature and strain rate. This is called adiabatic process, thus, it results in adiabatic stress $\sigma_{\text{adiabatic}}$. This information cannot be applied further analysis since it there is still temperature effect involved. The final goal is to identify the stress at constant temperature or so-called isothermal stress $\sigma_{\text{isothermal}}$.
There is an analytical approach to decouple them from each other using the following equation:

$$\sigma_{\text{adiabatic}} = \sigma_{\text{isothermal}} \cdot f(T)$$

where $f(T)$ is the thermal-induced softening function at the temperature T. This function requires expert to define an accurate description. Hence, the prediction model in this project will allow us to predict isothermal stress without help from the domain expert.
For this type of problem, there is a work from Liu et al. that took an attempt to model the elastic modulus and the yield strength by digging deep into microstructure level. Versino et al. presented their work on flow curve prediction of plastic deformation of a copper. Their environment was more extreme than us (temperature of 300 K and 800 K and strain rates up to 10^{12} s^{-1}) since it is a metal forming process.

Metrics

Since it is a regression problem, Mean-Squared Error (MSE) and r2 score are chosen as the metrics. Their calculations read:

$$\text{MSE} = \frac{\sum_{i=1}^{N} (\sigma_{\text{pred}_i} - \sigma_{\text{exp}_i})^2}{N}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{N} (\sigma_{\text{pred}_i} - \sigma_{\text{exp}_i})^2)}{\sum_{i=1}^{N} (\sigma_{\text{average}} - \sigma_{\text{exp}_i})^2)}$$

where σ_{pred} is predicted stress, σ_{exp} is stress measured from measurement, σ_{average} and N is number of data points. At the testing stage, the σ_{exp} becomes the pre-calculated isothermal stress at the strain rate of 1 x 10^{-1} and 1 x 10^{-2} s^{-1} by analytical approach.

Analysis

Data Exploration

The data set for this project is structured data. They are flow curves obtained from different temperatures and strain rates in adiabatic condition as well as analytical flow curve in isothermal condition for high speed tensile tests. The material in this project is a corrosion resistant steel, AISI 439. Its chemical composition in percentage mass is listed in Table 1.

C	Cr	Ti	Ni	Mn	Si	S	P
<0.05	16.0-18.0	0.15-0.80	<0.30	<1.00	<1.00	<0.015	<0.04

Table 1: Chemical composition of AISI 439

All the tests were done at our lab in the RWTH-Aachen University using a universal tensile test machine Zwick 100kN according to ASTM standard. The applied load and elongation were measured by a load cell and a strain gauge extensometer. The load and displacement information was transformed into stress and strain by the machine's built-in software. The strain rate can be exported to the data set easily since it is a controlled variable. The test temperature is controlled by applying the test in temperature chamber. However, a temperature sensor is also attached to the sample in order to evaluate the material's actual temperature. Therefore, the overall recorded information in the data set is listed in Table 2.

Strain, -	Temperature, K	Rate, s^{-1}	Stress, MPa

Table 2: Data set column name

As our goal is to predict the stress at a given environment information, **the stress become our label and the remaining are defined as features.** Different environment settings are designed to identify the material's response in different manners. The design of experiment matrix is shown in Table 3.

Temperature, K	Strain rates, s^{-1}				
	1×10^{-4}	2.78×10^{-4} (QS)	1×10^{-3}	1×10^{-2}	1×10^{-1}
293 (RT)	✓	✓	✓	✓Δ	✓Δ
373		✓		x	
473		✓			
573		✓			
673		✓			

Table 3: Design of experiment of tensile test at various conditions

The environments which are indicated by ✓ have experimental results in adiabatic condition. One the other hand, the ones which are also indicated by Δ yield calculated results from the analytical solution. **The** x **indicates a new data set that is included after the first submission. It has never been considered before. But there is one comment mentioned that the prediction model shall be tested its robustness. This will play a role as a robustness test set.** It can be noticed that only the cases with high strain rate shall separate the adiabatic condition and isothermal condition, the slower loading cases do not significantly induce thermal-softening effect.

Exploratory Visualization

The visualization of this project shall start from the flow curve of AISI 439 at RT and QS condition. This curve provides the characteristic of the material as how much it deforms if there is load applied at the very normal situation. The curve is illustrated in Fig. 2.

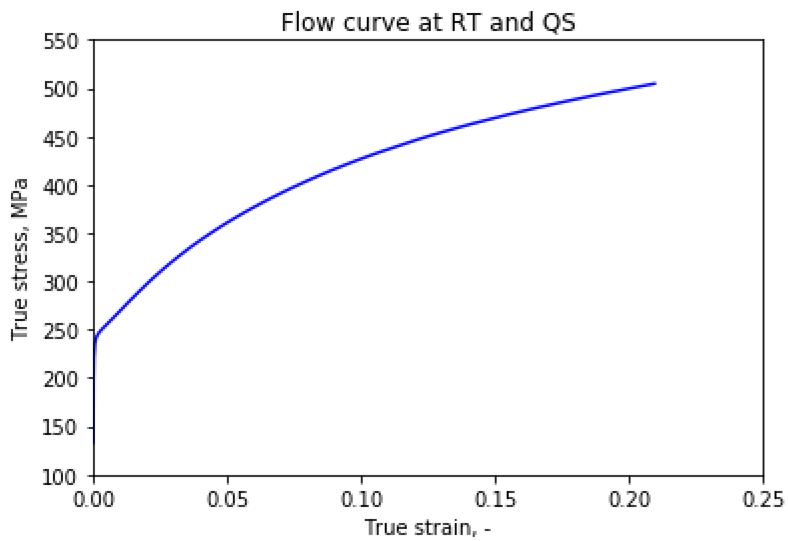

Figure 2: Flow curve of AISI 439 at room temperature (RT) and quasi-static (QS) loading speed

Tensile test results at elevated temperatures and high speed loadings are illustrated in Fig. 3 and Fig. 4 respectively. It can be noticed that once the temperature rises, the flow stress becomes lower. In other words, the material is softer. In contrast, if the loading speed become higher, the material become harder. This effect is known as strain-rate hardening effect.

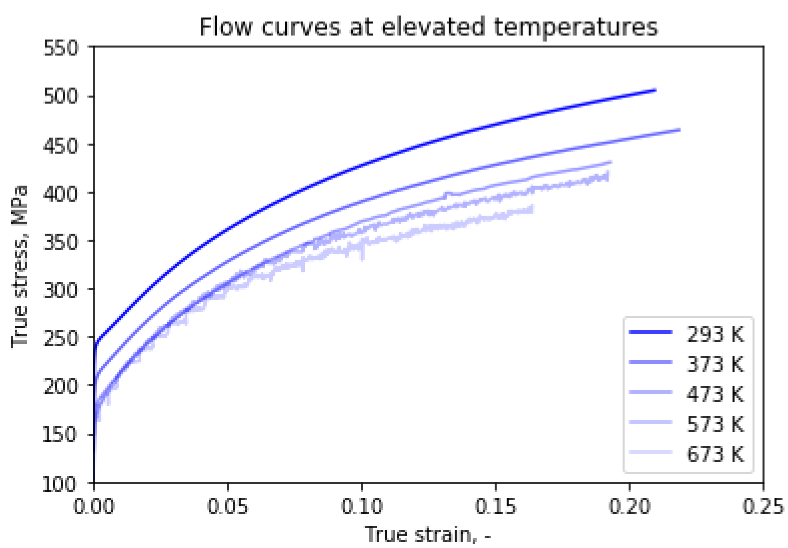

Figure 3: Flow curves of AISI 439 at various temperatures and QS loading speed

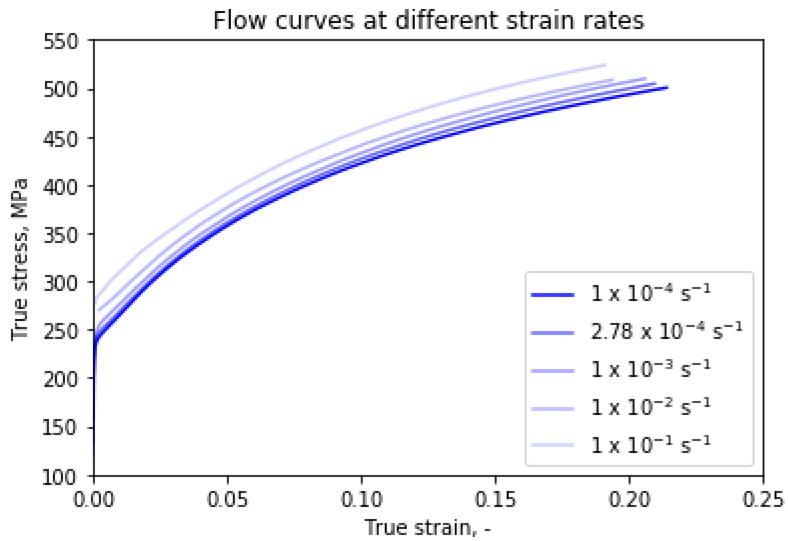

Figure 4: Flow curves of AISI 439 at various strain rates and at RT

At high speed loading conditions, if the speed is high enough, the heat dissipation rate is not high enough to allow heat to transfer to the material's ambient. From Fig. 5, it can be noticed that all tests start at RT $T = 293K$. As strain evolves, the temperature in the conditions $\dot{\varepsilon}$ = 1 x 10^{-1} s^{-1} and 1 x 10^{-2} s^{-1} rises up.

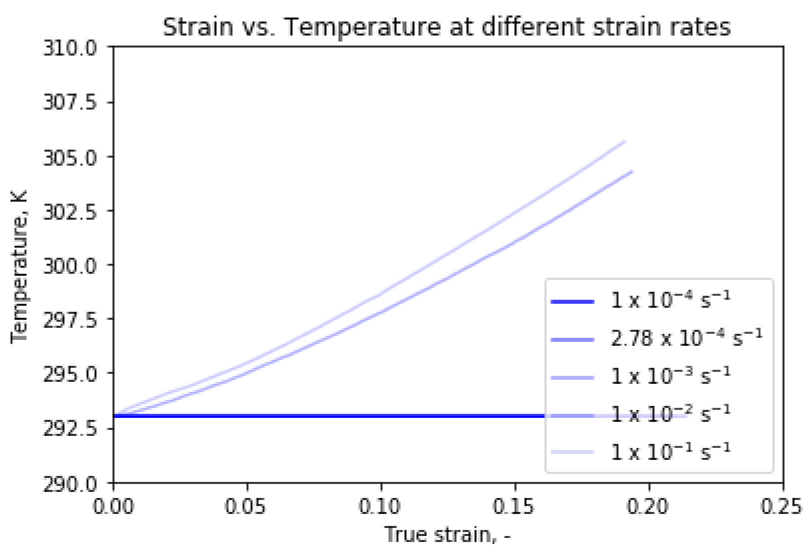

Figure 5: Temperature evolution during high speed tensile tests

The remaining feature, strain rate, does not need to visualize or identify any descriptive statistics since it is a controlled variable of the machine. It keep constant along its corresponding flow curve.

Algorithms and Techniques

As the nature of this problem is to predict stress based on given strain, temperature, and strain rate, it can be considered as a regression problem based on multivariate time-series analysis. This project proposes three different machine learning prediction algorithms which are:

1. Support Vector Regression (SVR)
2. Multi-level Perceptron (MLP) or Artificial Neural Network (ANN)
3. Long-Short Term Memory (LSTM)

This strategy allows us to test model ranging from a very simple to advance algorithm in time-series prediction. Each of which is taken into a hyperparameter optimization scheme. The SVR can easily applies the RandomizedSearchCV to find the optimal C, epsilon, and gamma. The MLP and LSTM require some additional wrapper from sklearn for the optimization scheme. The parameters taken into account for them are number of hidden layers and number of neuron in each layer.

To describe the first and foremost model, SVR, it should mention a linear regression algorithm first. It tries to form a line which minimize the error rate of every point. The SVR, on the other hand, forms an arbitrary line depending on the kernel type and allows some error but let those error inside a band.

The MLP or ANN functions like a human brain that construct multiple layers of neurons. Each layer take information from the previous one, calculate it, and send it again to the next layer. This does not include any memory inside. Thus, Recurrent Neural Network (RNN), the parent model of LSTM, was introduced. The LSTM's idea is like ANN but it includes memory inside so the model does have some idea about the current data based on what it sees before.

Benchmark

The prediction algorithms are benchmarked by comparing to analytical solution which is used by expert in this field to map adiabatic stress to isothermal stress. As mentioned before that in order to obtain such solution, there are several steps which require expert's decision. The function reads:

$$f(T) = C_1 \cdot exp(C_2 \cdot T) + C_3$$

where C_1, C_2, and C_3 are parameters to be calibrated. From the given experimental results, the calibrated values are listed as follows:

Parameter	Value
C_1	3.448
C_2	0.008
C_3	0.684

Table 4: Calibrated parameters for analytical solution of thermal-induced softening function

The flow curves at strain rate 1 x 10^{-1} s^{-1} and 1 x 10^{-2} s^{-1} with this function applied as isothermal stress are included in the given data set. They play a role as validation data set.

Methodology

Data Preprocessing

Data preprocessing in this project is composed of multiple steps as follows:

1. Apply decadic logarithm to strain and strain rate
2. Apply scaling scheme to all features to avoid ill-conditioned calculation
3. Include features in the previous time-step as the current time-step's feature

Due to the fact that the strain and strain rate are features that contain very low number, it would be a great idea to apply a logarithmic function to scale them down. Afterwards, all the features are then scaled again by a predefined range. This will allow values like stress and temperature which are in magnitude of hundreds to be in magnitude of less than one. In addition, the raw data set contains only 4 features in total. It might not be sufficient for the algorithms to make an accurate prediction. Therefore, including features in the previous time-step enables the models understand the how each variable develops as loading proceeds. At the current stage, it is found that including features in two previous steps yielding quite good results comparing to the training time.

The difficulty for this project first of all is the data extraction. It shall distinguish the adiabatic flow curves out of isothermal flow curves. However, at the early stage of the test in adiabatic condition, the temperature does not rise much. It will look like isothermal flow curves and in Pandas one might mispicking the partial flow curve. Using LSTM is another challenge to overcome. It has never been mentioned in the course content. Therefore, it would be better if there is topic can cover in the Machine Learning Nanodegree contents.

Implementation

This section dedicates to describe the approaches in applying each algorithms to separately. In fact, at the very first algorithms SVR yielding a satisfactory result. But it is challenging to implement a more complex model in order to learn along the way. Therefore, the implementation of a simple model like SVR is introduced first. Then the ANN shall be taken into account. Lastly, the widely used model for time-series analysis, LSTM, is considered. For this section, only Vanilla LSTM shall be mentioned.

SVR

The steps for SVR is quite straightforward. It consists of standard Scikit-learn workflow which are:

- Create a model
- Train model with training data
- Evaluate the model's performance, in this case negative mean squared error
- Validate the model by evaluating the model's performance using the isothermal flow curves at strain rates are 1×10^{-1} and $1 \times 10^{-2}\ s^{-1}$ as testing data
- The predictions on testing data are finally evaluated by r2 score

ANN

The implementation of ANN is mainly based on Keras library. In order to apply the model_selection module like *RandomizedSearchCV* or *GridSearchCV*, the model shall be wrapped by KerasRegressor function in Scikit-learn. The implementation process is identical to the SVR model which are create, train, evaluate performance on training data, validate on testing set, and calculate the r2 score to see how good is it.

LSTM

LSTM also use Keras as its main library for the implementation. In the same manner as ANN, the KerasRegressor plays a role as the wrapper of LSTM to make use of Scikit-learn function. The LSTM in this project is Vanilla LSTM which has only one LSTM layer and a fully-connected layer.

Refinement

At first the models are evaluated using default parameter set. In the refinement section, each of them use the *RandomizedSearchCV* to identify the best hyper-parameter set from given distribution. The parameter set and its corresponding distribution for each model is listed in Table 5.

Model	Hyper-parameter	Distribution
SVR	C Gamma epsilon	stats.uniform(1000, 10000) stats.uniform(0.00001, 0.1) stats.uniform(0.0001, 0.1)
ANN	Number of hidden layers Number of neuron in each layer	np.random.randint(0, 10) np.random.randint(1, 20, size=n_layers)
LSTM	Number of unit in LSTM layer	range(1, 200, 1)

Table 5: Model hyper-parameter list and their corresponding distributions

In this particular case, the *RandomizedSearchCV* does not work properly in distribution of ANN. Thus, *GridSearchCV* is implemented instead. While the parameters selection is based on random function, it does not work differently.

Results

Model Evaluation and Validation

The first evaluation is the r2 score during the training phase. The values are based on the best hyper-parameter set. The score for each model can be found in Table 6

Model	r2 score during training phase
SVR	0.999564
ANN	0.998809
LSTM	0.999929

Table 6: Training r2 score for each model

These models are validated by applying to the isothermal conditions to identify the isothermal flow curve. The prediction results for SVR, ANN and Vanilla LSTM are illustrated in Fig. 6, Fig. 7, and Fig. 8 respectively. The r2 score of each model is listed in Table 7.

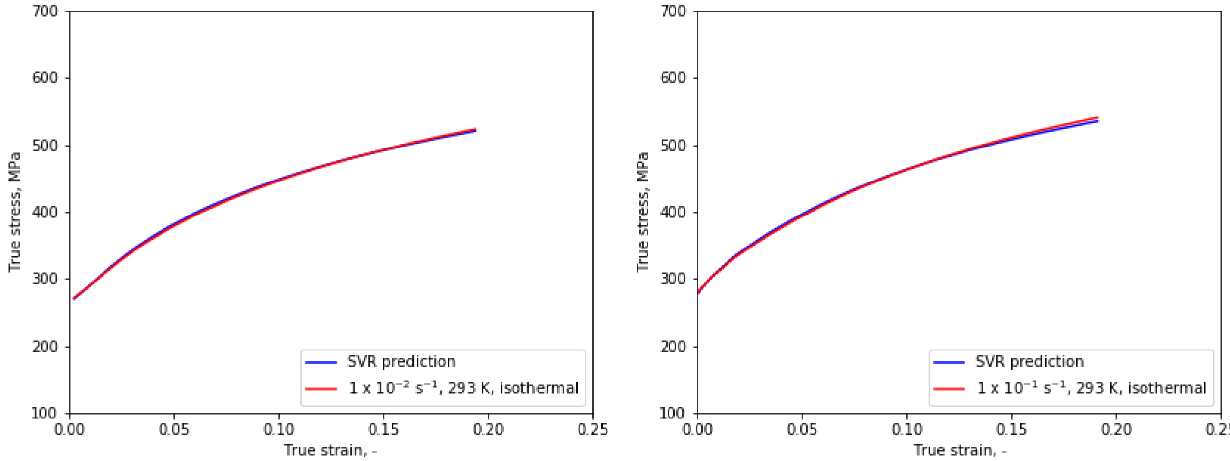

Figure 6: Comparison between SVR prediction vs. analytical isothermal flow curve (left) at 1 x 10-2 s-1 strain rate (right) at 1 x 10^{-1} s^{-1} strain rate

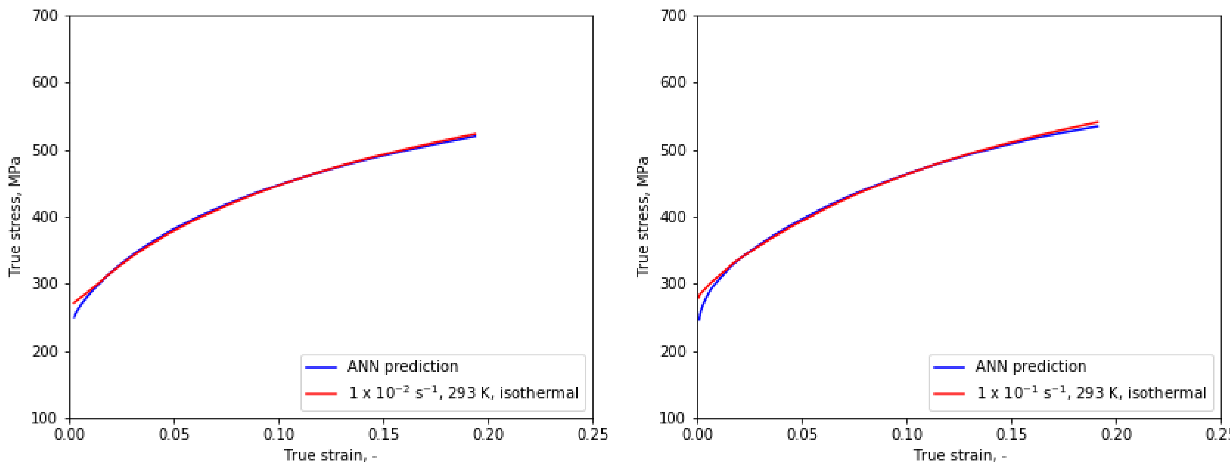

Figure 7: Comparison between SVR prediction vs. analytical isothermal flow curve (left) at 1 x 10-2 s-1 strain rate (right) at 1 x 10^{-1} s^{-1} strain rate

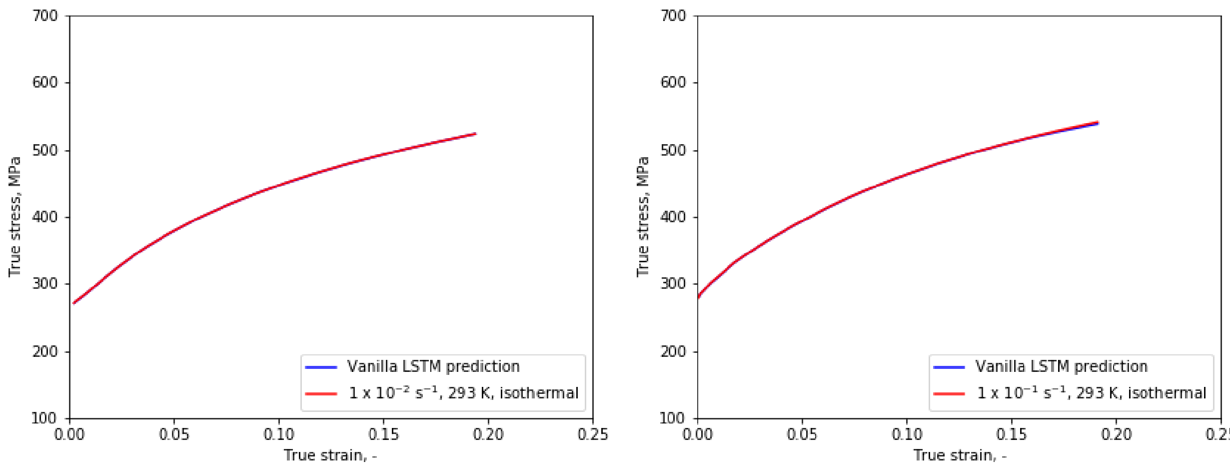

Figure 8: Comparison between SVR prediction vs. analytical isothermal flow curve (left) at 1 x 10-2 s-1 strain rate (right) at 1 x 10^{-1} s^{-1} strain rate

Model	r2 score for 1 x 10^{-2} s^{-1} strain rate test set	r2 score for 1 x 10^{-1} s^{-1} strain rate test set
SVR	0.999303	0.998667
ANN	0.998150	0.996332
LSTM	0.999988	0.999878

Table 7: Testing r2 score for each model

The models shall be tested again with an unknown set which has just been introduced after the first project submission. The results for SVR, ANN, and LSTM are shown in Fig. 9, Fig. 10, and Fig. 11 respectively. In the same manner, the r2 score for each model is listed in Table 8. It can be noticed that the ANN underperforms as usual and LSTM still keeps its best relative performance.

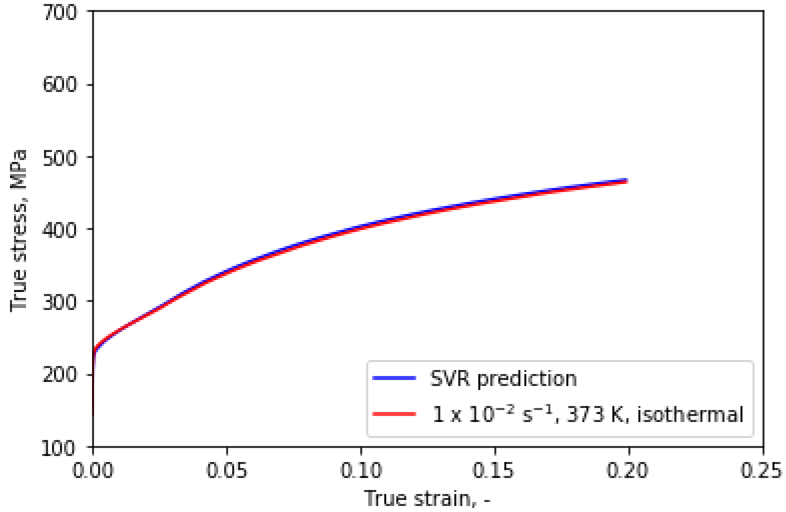

Figure 9: Comparison between SVR prediction and unknown dataset

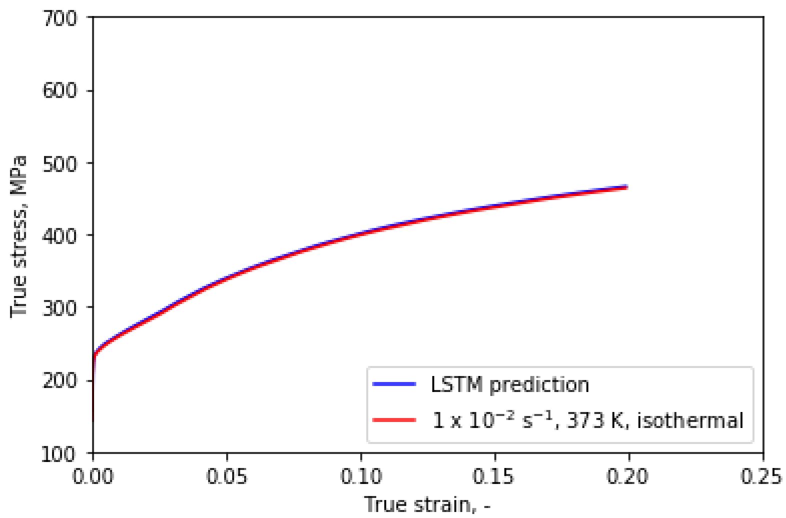

Figure 10: Comparison between ANN prediction and unknown dataset

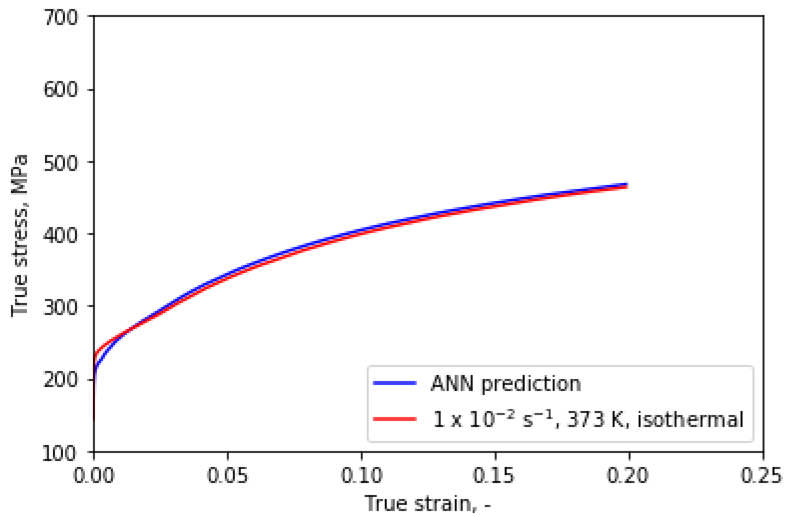

Figure 11: Comparison between LSTM prediction and unknown dataset

Model	r2 score for unknown data set
SVR	0.998133
ANN	0.990678
LSTM	0.998957

Table 8: Testing r2 score with unknown data set

Justification

It is quite satisfactory that the models can predict the testing set in the r2 score above 99.5%. Nevertheless, there are two points to be mentioned in the results:

1. LSTM yields the best score on the testing set. It outperforms other models as expected since LSTM take an advantage from its memory and forget gate.
2. SVR, on the other hand, performs better results than ANN in both training and testing set. It could be due to the fact that the behavior of the material is not much complicated. Thus, SVR can achieve sufficiently accurate results at some extent. But at the end, LSTM obtains the best results in testing set.

Conclusion

Let us recap about this project. It starts from collecting information from experiments which are in different conditions like temperatures and strain rates. Most of the tests were done on rather standard setting, either at room temperature or at quasi-static

loading. There is one data set that performed in high speed and high strain rate. It was considered as the robustness testing data set. Next, the collected data shall be preprocessed by the following procedure:

- **apply decadic logarithm to strain and strain rate**
- **apply scaling scheme to avoid ill-conditioned calculation**
- **include features in the previous time step as the current time step's feature**

The after the features and label are consistent among the flow curves. The training and testing set shall be distinguished to let the model train on the training data set. Three models, SVR, ANN, and LSTM were considered as our candidate models. Their best hyper-parameters were identified by *RandomizedSearchCV* except the ANN that use *GridSearchCV*. But the hyper-parameters in the grid are based on random function. Thus, it can still be considered as a random search. The models with their best hyper-parameters are then tried to predict the flow curve in isothermal condition. They performed quite well such that the r2 score are always over 0.99. The score can be sorted from highest to lowest by, ANN, SVR, and LSTM. At the end, the robustness testing with an unknown data set is performed. With this data, the ranking still do not change. That ensure the performance of LSTM that is widely used in time-series analysis.

Free-form Visualization

The r2 scores in overall are plotted in Fig. 12. SVR can achieve good r2 score in training set. Its predictions in both testing sets are lower but not much significant. In comparison, ANN yield not much lower r2 score in training and the difference between training and testing are quite obvious. It seems to be a little bit overfitting. And the last model, LSTM, yield the best results in every cases.

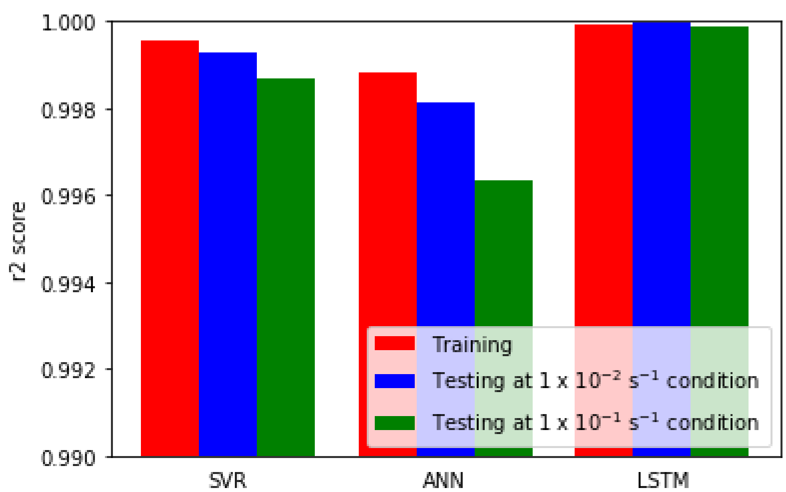

Figure 12: Comparison between training set and testing at different conditions for each prediction model

Reflection

This project allows me to think about machine learning process in regression problem. There are few points that were not considered during the course since most of them are prepared. For this one, the problem should be thought and prepared thoroughly before submitting the proposal.
At first, the detail about LSTM was not in my mind at all. What I knew was only “LSTM is the prediction model that people use for time-series analysis”. The contents were not covered in the course. But the this problem is something that should be solved in everyday life for people in material mechanics field. Therefore, I selected it as my capstone project. During proposal writing I still did not know that there are subtypes of LSTM like Vanilla LSTM, Stacked LSTM, etc. This is why I did not include the architecture of the model in the proposal. After accepted, I tried to find some existing materials and code examples on the internet. That helped me until this project is finished.

Improvement

It is still interesting if the model can perform this well if there is no stress information in any time-step available in the features. The stress values, even if they are from previous time-step, seem to be a key features for the prediction.
In case if, the prediction models mentioned above cannot perform well once the stress values are excluded from the features. It might be interesting how well the other types of LSTM can perform.