NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

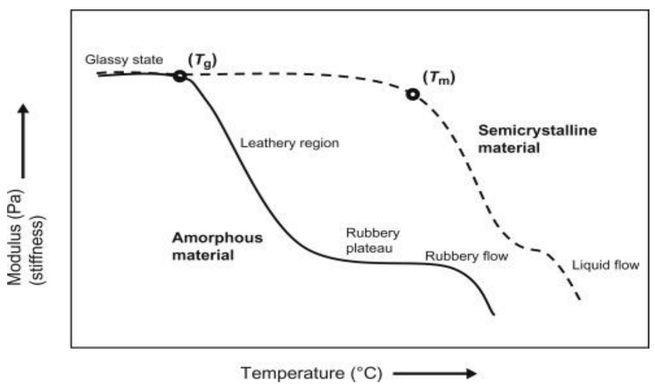
Odyssey Research Programme
School of Physical and Mathematical Sciences

Predicting Glass Transition Temperatures of Polymers

Goh Kai Leong

Supervised by Dr Lu Yunpeng* and Asst Prof Xia Kelin

Background & Objective



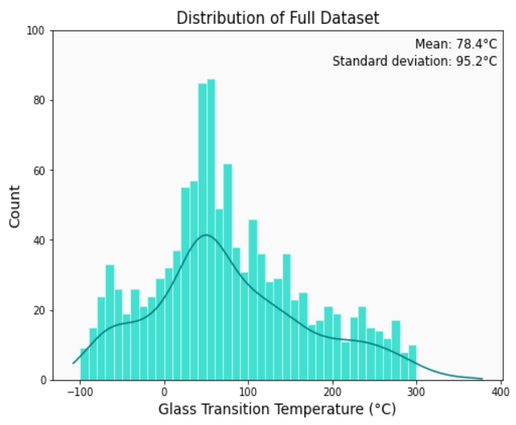
Glass transition temperature (T_g) is the temperature at which a polymer changes from a glassy state to a rubbery state. This drastic change of property in polymers makes T_g very important in food processing and drug synthesis. [1]

The objective is to develop a regression machine learning model that can accurately predict T_g.

Methodology

Dataset

- Dataset of 1200 polymers was compiled from PolyInfo database
- Polymers consist of H, C, N, O, F, Si, P, S, Cl, Ge, Se, Br (12 non-metal elements)
- The T_g values were found to be roughly normally distributed



Molecular Geometry Optimization

Using chemistry-related software like GaussView, Gaussian and Avogadro to edit and optimize the molecular data [2]

Feature Selection

- Mordred library (2D + 3D molecular descriptors) [3]
- RDKit library (Extended connectivity fingerprints) [4]
- Subsequently combine the features and remove redundant bits

Machine Learning Algorithm

Extreme Gradient Boosting Regression [5]
An enhanced form of Gradient Boosting
Works by combining weaker decision trees to form a single stronger model

Algorithm evaluation

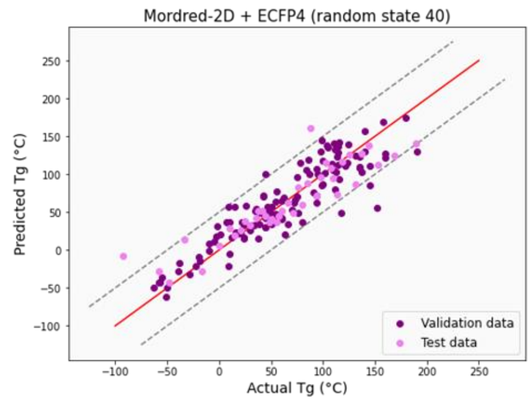
- Coefficient of determination (r^2) - Higher is better
- Root-mean-square error (RMSE) - Lower is better
- Cross-validation (CV) – To check for any overfitting

Results

Previous Work

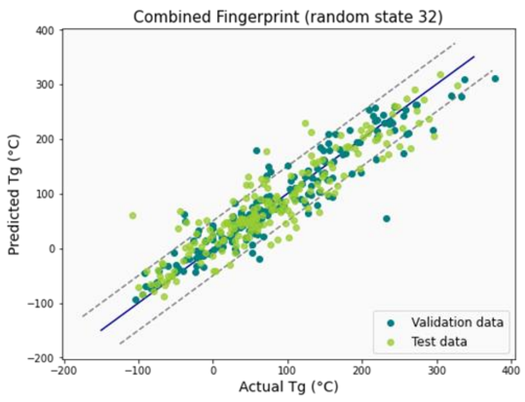
- Comprised 640 data points
- Polymer data from Polyinfo
- Similar methodology applied

Evaluation	r^2	RMSE
Validation	0.813	24.5
Test	0.812	26.9
CV	0.712	32.6



Current Work

- Comprised 1200 data points
- Polymer data from Polyinfo
- r^2 improved but RMSE got worse



Evaluation	r^2	RMSE
Validation	0.881	33.2
Test	0.844	37.5
CV	0.841	38.1

Conclusion & Future Work

With the increase in number of data points:
- The r^2 improved while RMSE became worse

In the later stage of our research, we intend to try:
- Bringing r^2 closer to 1
- Bringing RMSE lower
- Further increasing the number of data
- Consider deep learning

References

1. Shrivastava, A. Introduction to Plastics Engineering. In *Introduction to Plastics Engineering*; William Andrew, **2018**; pp 1-16. https://doi.org/10.1016/B978-0-323-39500-7.00001-0
2. https://gaussian.com/citation/ & https://avogadro.cc/cite/
3. Landrum, G. An overview of the RDKit. http://www.rdkit.org/docs/Overview.html (accessed May 12, 2020)
4. Moriwaki, H.; Tian, Y.; Kawashita, N.; Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminformatics*. **2018**, *10*, 4
5. XGBoost Developers. XGBoost Tutorials. Introduction to Boosted Trees. https://xgboost.readthedocs.io/en/latest/tutorials/model.html (accessed May 12, 2020)