Imperial College London

Department of Earth Science and Engineering

MSc in Applied Computational Science and Engineering

Independent Research Project

Project Plan

Accelerating flash calculation using Machine Learning

by

Hanlin XIA

Email: dombledore.Xia21@imperial.ac.uk
GitHub username: acse-dx21
Repository: https://github.com/acse-dx21/IRP_flash_calculation_accelaration.git

Supervisors:
Dr. Salinas, Pablo
Professor Jackson, Matthew D
Santos Silva, Vinicius

July/September 2022

Introduction

Phase equilibrium (flash) calculations have been a long-standing hot topic in thermodynamics. Using method such as Successive substitution method to correlate and predict vapor-liquid equilibrium takes a lot of time due to long-term iteration.

Because of the high time consumption in traditional way, attention have been focused on other model and try to find better solutions. Yu Li et al (2019) introduced a new model through deep neural network (DNN) method trying to replace this process. Deep neural network, an important structure in machine learning methods, have had a great impact in different fields, including, self-driving cars, identification of images. Due to its black-box nature and flexibility, it has been being used in different fields to try to solve long standing issues with great success in some areas. Yu et al 2019, using variations of DNN and one can obtain very good results for flash calculations. It is even better than Newton's method and the Sparse Grids Method in some respects such as efficiency to implement. While during the same era, physical informed neural networks (PINNs) was introduced and some of them are shown to have good performance in some situation. (Thelma Anizia Ihunde,2022)

Despite of deep learning, ensemble-based methods have also been used in this region. In 2016, T. Chen et al made a study showed that traditional machine learning methods (like Gradient boosting) had dominant status in processing tabulate data. Also in 2022, Ravid Shwartz-Ziv et al give a wide range performance test between deep learning model and tree ensemble model (such as XGBoost) in processing tabulate data. They conclude that XGBoost outperforms these deep models across the datasets. Besides, when neural network meets reality problem, it is often accompanied by a range of thorny problems such as lack of locality, data sparsity and mixed feature types since the input need to be well defined. While at the same time, tree ensemble-based methods often require much less time tuning and less computation cost to fit data. At the end of the paper, Ravid et al (2022) shows that combination of deep models and XGBoost can perform better on these datasets than single model alone

Due to the success of DNN for flash calculations, but simultaneously the known problems to deal with tabulated data, in this project we are going to study the use of ensemble-based methods for flash calculations and compare its performance with DNNs. Moreover, we will study the effect of PINNs and the long-term effect of performing several consecutive predictions.

Literature Review

Vapor-liquid equilibrium (VLE) calculation is of great importance in modeling and simulating flows with multiphase and multicomponent (Shuyu Sun et al ,2007; T.C. Tan et al, 2004). The method to calculate equilibrium information has been studied by scientist for a long time. In the last few decades, people use equations of state method (EOS) to solve this problem (Yu Li et al,2019). The EOS method, to simulate a real process can produce accuracy result. However, in most cases, such method needs a lot of time for iteration, which often limits the practicability in real scenarios. To deal with this problem, several

methods several methods have been implemented such as Newton's method and Sparse grids method and both have great performance.

In 2012, the breakthrough of AlexNet draw people's attention into deep learning model. The model, with great ability, showed an excellent performance for a wide variety of applications (Yu Li et al,2019). Yu Li et al 2019 studied how to implement artificial neural networks in VLE calculation and compare it with traditional algorithms. They found that deep learning does have advantages in VLE calculation and can even outperform some well-accepted algorithms in processing such tabulate data such as support vector machine (SVM), which perform bad in large dataset.

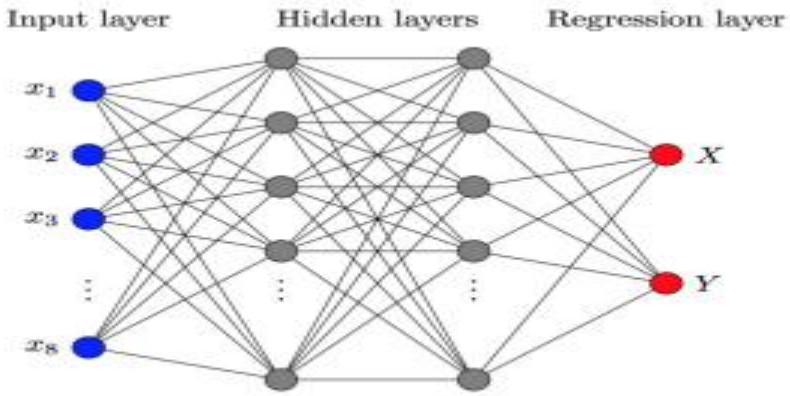

Fig. 1. The neural network to model VLE

The ANN Li et al 2019 used, like other neural network, takes several variables as input such as critical pressure (P c), critical temperature (T c) and acentric factor(ω) of each component in mixture. The model process input data using linear combination layer and activation function, then the output is transferred into next layer and so on. With back propagation used in the neural network, it can decrease the gap between prediction and true values automatically.

While neural networks by Li et al achieve great success in efficiency, the fix number components requirement remains a problem. Tao et al 2020 introduced a model of two network structures with self-adaptive mechanism attempt solve this problem in some extent. And as respect to practical gas-liquid conversion, some results generated by neural network are not realistic due to the physical constraints such as mass balance. Thelma et al 2022 show that by add physical constrained in loss function, the neural network can reduce physics constraint errors and thus more realistic and useful.

In 2022, a new paper written by Ravid Shwartz-Ziv et al, shows that Tree ensemble models (like GDBT) are usually outperform deep model in processing tabulated data despite of great development deep learning yet. In this paper, they did a thorough study by exploring different applications' (such as DNF-Net, XGBoost) performance in different dataset (such as Microsoft, Shrutime) and systematically compare the models' behavior in various situation. In the end of the paper, the authors concluded that XGBoost was able to outperform this deep models. Additionally, they conclude that an ensemble of XGBoost and deep models can obtain better results than XGBoost alone.

Description of Problem and Objectives

VLE are very costly to do and introduce instabilities into the system. The algorism with too much iteration is apparently not the best algorism to implement. Thus, surrogate model is required in this place.

There has been already an attempt to substitute forward models using ANN with excellent results (Yu Li et al 2019). However, the authors did not explore other type of machine learning approaches and neither they tested PINNs nor the effect of a potential accumulated error from continuously using ML to perform the VLE operations. Whereas deep learning in physics is still relatively new, the model should have a lot of room for improvement.

The objectives of this project are:

1) Test different type of ML methods that have been shown to work better for this scenario, including tree ensemble model and deep model

2) Test the use of PINNs/add physical constrain to predictions. Test if they can generate better result

3) Study the effect of the accumulated error when using ML methods to continuously predict VLE.

Progress to Date and Future Plan

Dates	**Task**
6 Jun – 10 Jun	Literature review and build Generate data library. -----Done
10 Jun – 20 Jun	Form test module/log module/simple model (GBDT, ANN) to build more powerful and stable library -----Done
20 Jun – 30 Jun	Test the simple model and try to fix it into larger dataset and build system that used to find best hyperparameter of model -----Done
1 July-31July	implement more models with best hyperparameter and collect data to analysis it, this contains 1 use more model, like Random Forest 2 use model trained with more information, like PINNS/physical constrains And check performance of this application.
1 July- 10 July	visualize and analysis data, and study repetitions
After 10 July	Write report, final check of the research

Reference

Ravid Shwartz-Ziv, Amitai Armon, Tabular data: Deep learning is not all you need, Information Fusion, Volume 81,2022, Pages 84-90, ISSN 1566-2535,
URLhttps://doi.org/10.1016/j.inffus.2021.11.011.

Shuyu Sun, Mary F. Wheeler, Discontinuous Galerkin methods for simulating bioreactive transport of viruses in porous media,Advances in Water Resources,Volume 30, Issues 6–7,2007,Pages 1696-1710,ISSN 0309-1708,
URL https://doi.org/10.1016/j.advwatres.2006.05.033.

T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

T.C. Tan, C.M. Chai, A.T. Tok, K.W. Ho, Prediction and experimental verification of the salt effect on the vapour–liquid equilibrium of water–ethanol–2-propanol mixture, Fluid Phase Equilibria,Volume 218, Issue 1,2004,Pages 113-121,ISSN 0378-3812,
URL https://doi.org/10.1016/j.fluid.2003.11.010.

Tao Zhang, Yu Li, Yiteng Li, Shuyu Sun, Xin Gao,A self-adaptive deep learning algorithm for accelerating multi-component flash calculation,Computer Methods in Applied Mechanics and Engineering,Volume 369,2020,113207,ISSN 0045-7825,
URL https://doi.org/10.1016/j.cma.2020.113207.

Thelma Anizia Ihunde, Olufemi Olorode, Application of physics informed neural networks to compositional modeling, Journal of Petroleum Science and Engineering, Volume 211, 2022, 110175, ISSN 0920-4105,
URL https://doi.org/10.1016/j.petrol.2022.110175.

Yu Li, Tao Zhang, Shuyu Sun, Xin Gao, Accelerating flash calculation through deep learning methods, *Journal of Computational Physics*, Volume 394, 2019, Pages 153-165, ISSN 0021-9991.
URL https://doi.org/10.1016/j.jcp.2019.05.028.