**Introduction**

Accelerating flash calculations have been a long-standing hot topic in thermodynamics. Using traditional algorism such as Successive substitution method to correlate and predict vapor-liquid equilibrium have met a great difficulty in time consumption since the traditional way often need long time iteration to reach the target statement.

Due to the main drawback 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 learning method trying to replace this process. The deep learning method, also called neural network, have been shown to great impact in tremendously range of region including industry and the academia. Due to far exceeding other model’s accuracy and broad versatility in all kinds of region, deep leaning method has become one of the most popular models in the past few years. The paper by Yu, also using the families of this model and show good result in flash calculations, even by cross comparation to other model, such as Newton’s method and the Sparse Grids Method.

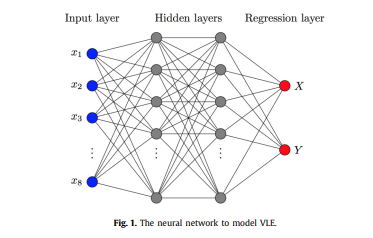
However, In 2022, Ravid Shwartz-Ziv(et al) introduced another family of model named Tree ensemble models, which also shows the great power of models other than neural network. The Tree ensemble model family (such as XGBoost and gradient-boosted decision trees (GBDT)) still shows the dominant status in processing tabulate data. 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. While at the same time, the XGBoost requires much less tuning and less computation cost to fit data. In addition, Ravid shows that ensemble of deep models and XGBoost performs better on these datasets than single model alone

The main goal of this study is to find a model to accelerate flash calculation, while at the same time, we need to compare the existing models mentioned in the two paper above and try to Evaluate in different dimensions. While also we can also build library to help us implement this method in further study.

**Literature Review**

Vapor-liquid equilibrium (VLE) calculation is of great importance in modeling and simulating flows with multiphase and multicomponent. In the last few decades, people raised numerous experiences and obtain tremendous data trying to enhance model performance with equations of state method (EOS). The EOS method, simulate real process in experience like combination and interaction between binary mixtures, can produce accuracy result in the end. However, in most cases, such method needs a number of iterations to reach the equilibrium statement, which often limits the practicability in real case. In order to deal with this problem, several methods were raised such as Newton’s method and Sparse grids method. Both of them did well in their age.

In 2012, the breakthrough of AlexNet draw people’s attention into deep learning model. The model, with amazing ability, showing its strength in all kinds of region. It can not only carry out well in original computer vision field, setting new record in image recognition and language translation, but can also been noticed to have great performance in many other problems such as protein binding affinity prediction [5], enzyme function prediction [17], structure super-resolution reconstruction and modeling brain circuits (Yu Li et al,2019). Yu Li(2019) studied how to implement artificial neural networks in VLE calculation and compare it with traditional algorithms. He found out that deep learning does have advantage in VLE alculation and even outperform some “elder brother” algorithms in processing such tabulate data such as support vector machine (SVM)



The ANN Li used, similar to other neural network, takes several variables as input, such as critical pressure (P c), critical temperature (T c) and process it by linear combination layer and activation function, Then it is transfered into next layer and so on, the math formular will look like this:

yi = fi(Wi ∗ ai + bi),

where Wi is the weight; bi is the bias; and fi is the activation functions of the i-th layer

Meanwhile, He also use multiple technique like Weight decay and Dropout method to gain better function of model

It seems that deep learning have become Silver bullet in all kind of region, however in 2022, a new paper writed by Ravid Shwartz-Ziv, Reestablished leadership of machine learning method in processing tabulate data by using Tree ensemble models (such as XGBoost). The author also admits that some deep networks do have strength in some situation as well. In this paper, He collected all kinds of model (such as DNF-Net, 1D-CNN) and all kinds of dataset from various place (such as Microsoft, Shrutime) on the Internet to systematically compare the performance between models and between datasets. In the end of the paper, the author shows that XGboost perform better compare with other model and In addition he conclude that Ensemble model of XGBoost and deep models can score better results than XGBoost alone. The detail of study that Ravid have down will be introduced below:

**Differentiable trees**: Family member of old brothers in processing tabulate data and have excellent performance. It takes advantage of good property in Quadratic function and fix some training problems in classical decision trees. In addition, it supports parallal programming and approximate algorithms, which is really good in our practical study.

**Attention-based models:** Attention-based model has been widely used since 2017. Recent works have suggested that not only inter-sample attention, but intra-sample attention is also important.

**Model ensemble:** modeling ensemble is a widely accepted way to generating better result and reduce fluctuation in different dataset. Consequently, ensembles tend to increase it stability and get better results

And also, these models have been used in various dataset (like Gesture Phase, Gas Concentrations, Microsoft (MSLR) and so on). Several analysis methods have also been used like Statistical significance test. Due to his Scientific and Systematic work, the article gives an instructiveness suggestion in this region and in our future study.

**Description of Problem and** **Objectives**

-------to be continue

**Progress to Date and Future Plan**

|  |  |
| --- | --- |
| **Dates** | **Task** |
| 1 Jun – 10 Jun | Literature review and build Generate data library. |
| 10 Jun – 20 Jun | Form test module/log module/simple model(XGB,ANN) to build more powerful and stable library |
| 20 Jun – 30 Jun | Test the simple model and try to fix it into larger dataset and build system that used to find best parameter of model |
| 1 July-20 July | Try to implement more complex model and collect data to analysis it |
| 20 July- 25 July | visualize and analysis data, revice what need to be improved |
| After 25July | ## Haven’t figure out, beside writing report, what can I done? |

**Reference**

-------to be continue