Neural Simulation Pipeline for Liquid State Machines

Karol Chlasta's PhD supervisor:
dr hab. Grzegorz Marcin Wójcik, prof. UMCS, prof. PJATK
Auxiliary PhD supervisor:
dr hab. Izabela Krejtz, prof. USWPS

Keywords: neuroinformatics, scientific workflows, containerisation, computer simulations, liquid state machine (LSM)

Author claims that numerical simulations must integrate a robust model development methodology, with adequate testing and simulation steering workflows to increase scientific throughput and improve utilisation of current and next-generation computational infrastructure, available both on-premise and in-cloud. To this end, there is the need to transform the end-to-end computational experiment workflow from one that is non-universal and manual to one that is standardised and (at least partially) automated.

Liquid State Machines (LSMs) are a type of recurrent neural network that have been widely used for tasks such as pattern recognition and classification. However, simulating LSMs can be computationally expensive due to their large number of neurons and connections. In this PhD thesis, author presents a novel Neural Simulation Pipeline (NSP) for LSMs that significantly reduces the computational cost of simulation while automating the tasks needed to manage and deploy the experiments into different execution environments. A provider-agnostic simulation framework can be used to run simulations on different hardware platforms or microprocessor architectures to allow researchers to use the most appropriate hardware and software for their specific simulation needs, without being tied to a specific vendor or provider. In the High Performance Computing (HPC) context, public cloud resources are becoming an alternative to the expensive on-premise clusters.

This thesis presents Neural Simulation Pipeline (NSP), a set of Bash and PowerShell scripts to facilitate the large scale computer simulations and their deployment to multiple computer infrastructures using the infrastructure as code (IaC) containerisation approach. The pipeline consists of three main components: a data preprocessing module, a simulation module, and a post-processing module. The preprocessing module manages the experiment's input data into a format suitable for LSM simulation, while the simulation module performs the actual experiment execution using a selected simulation engine. The post-processing module then analyses the simulated data and generates the final results.

Author demonstrates the effectiveness of NSP in a pattern recognition task programmed with GENESIS (a general purpose simulation engine for neural systems) and simulated through two custom-built visual systems: (1) RetNet(4x8,1) based on a single LSM column of multiple sizes, and (2) RetNet(28x28,4) using four LSM columns. Both systems were built using biologically plausible Hodgkin–Huxley spiking neurons. and are explored in the experimental chapters. The key finding relates to twelve different LSM readout algorithms, evaluated through five standard classification metrics, using the 10-fold cross-validation process. The LSM system presented achieves a repeatable accuracy and F1 Score of 81% for the readout based on Light Gradient Boosting Machine.

Moreover, the pipeline is evaluated by performing additional 54 experiments executed on-premise, and in the AWS Public Cloud environment. Author compares the standard and containerised execution, as well as presents the cost of execution in AWS. The results show that the NSP can significantly reduce entry barriers to LSM simulations, making it more practical and cost effective for real-world applications. The experimental conclusions are supplemented with practical tips related to prototyping with author's custom single board computer cluster (Neural Simulation Cluster), and suggested further research on the pipeline.