

## Potok symulacji neuronowej dla maszyn stanów płynowych

Promotorem mgr Karola Chlasty we wniosku o wszczęcie postępowania o nadanie stopnia doktora w dziedzinie nauk inżyniersko-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja jest dr hab. Grzegorz Marcin Wójcik, prof. UMCS, prof. PJATK. Jego promotorem pomocniczym jest dr hab. Izabela Krejtz, prof. USWPS

**Słowa kluczowe: neuroinformatyka, przepływy obliczeń naukowych, konteneryzacja, symulacje komputerowe, maszyna stanów płynowych (LSM)**

Autor twierdzi, że symulacje numeryczne powinny integrować metodologiczne podejście do rozwoju i testowania modeli cybernetycznych, z odpowiednimi przepływami pracy obliczeniowej. Uczynić to należy w celu zwiększenia ich skuteczności naukowej oraz poprawy wykorzystania obecnej i przyszłej infrastruktury obliczeniowej, dostępnej zarówno w modelu tradycyjnym, jak i w chmurowym środowisku obliczeniowym. W tym celu istnieje potrzeba przekształcenia przepływu pracy dla eksperymentów obliczeniowych z takiego, który jest nieuniwersalny i manualny na taki, który jest znormalizowany i (przynajmniej częściowo) zautomatyzowany.

Maszyny stanów płynowych (ang. Liquid State Machines, lub w skrócie LSM) są rodzajem rekurencyjnej sieci neuronowej, która jest szeroko stosowana w zadaniach takich jak rozpoznawanie i klasyfikacja wzorców. W niniejszej pracy autor przybliży tę tematykę i wskazuje, że symulacje takie mogą być kosztowne obliczeniowo ze względu na dużą liczbę wiarygodnych, kolczastych neuronów pulsacyjnych Hodgina-Huxleya i mnogość połączeń między nimi. W niniejszej pracy doktorskiej autor przedstawia nowatorski Potok Symulacji Neuronowej (ang. Neural Simulation Pipeline, lub w skrócie NSP) dla maszyn LSM, który znacząco zmniejsza koszt obliczeniowy takich symulacji, automatyzując jednocześnie zadania potrzebne do zarządzania i wdrażania eksperymentów obliczeniowych w różnych środowiskach uruchomieniowych. Potok umożliwia przeprowadzanie symulacji komputerowych w standaryzowany sposób na różnych platformach programowych i sprzętowych (w tym architekturach mikroprocesorowych), ułatwiając badaczom wykorzystanie najbardziej odpowiedniego sprzętu i oprogramowania dla ich specyficznych potrzeb symulacyjnych, bez przywiązania ich do konkretnego dostawcy usług lub architektury. W kontekście superobliczeń HPC (ang. High Performance Computing), zasoby chmury publicznej stają się alternatywą dla drogich klastrów komputerowych działających w tradycyjnych centrach obliczeniowych.

Wspomniany potok symulacji neuronowej NSP, będący zestawem skryptów Bash i PowerShell ułatwiających przeprowadzanie symulacji komputerowych na większą skalę i wdrażanie ich w różnych środowiskach uruchomieniowych korzysta z podejścia kontenerowego w paradygmacie Infrastruktura jako kod (ang. Infrastructure as Code, lub w skrócie IaC). Sam potok składa się z trzech głównych komponentów logicznych: modułu wstępnego przetwarzania danych, modułu symulacyjnego oraz modułu powykonaniowego. Moduł przetwarzania wstępnego zarządza danymi wejściowymi eksperymentu do formatu odpowiedniego dla symulacji LSM, podczas gdy moduł symulacji wykonuje rzeczywiste wykonanie eksperymentu przy użyciu wybranego silnika symulacyjnego. Moduł powykonaniowy wspomaga analizę danych symulacyjnych i generuje ostateczne wyniki.

Autor demonstruje skuteczność NSP w zadaniu rozpoznawania wzorców zaprogramowanym w GENESIS (silnik symulacyjny ogólnego przeznaczenia dla systemów neuronowych) i symulowanym przez dwa zbudowane przez siebie systemy wizyjne: (1) RetNet(8x5,1) oparty na pojedynczej kolumnie LSM o zmiennym rozmiarze oraz (2) RetNet(28x28,4) wykorzystującym cztery kolumny LSM. Oba systemy zostały zbudowane z wykorzystaniem biologicznie wiarygodnych neuronów impulsowych Hodgkina-Huxleya i są omówione w rozdziałach eksperymentalnych niniejszej rozprawy. Kluczowy wynik badawczy dla zaproponowanej architektury maszyn stanów płynowych dotyczy zbadania dwunastu różnych algorytmów warstwy odczytującej, ocenionych za pomocą pięciu standardowych metryk oceny klasyfikacji, z wykorzystaniem procesu 10-krotnej walidacji krzyżowej. Uzyskany przez autora wynik zaproponowanego systemu LSM osiąga dokładność (oraz F1 Score) na poziomie 81%, dla odczytu opartego o algorytm drzew decyzyjnych LightGBM (ang. Light Gradient Boosting Machine).

Ponadto, potok jest oceniany poprzez wykonanie 54 dodatkowych eksperymentów obliczeniowych wykonywanych w tradycyjnym centrum danych i w środowisku publicznej chmury obliczeniowej AWS. Autor porównuje standardowe i skonteneryzowane wykonania symulacji maszyn stanów płynowych, jak również przedstawia koszt wykonania poszczególnych eksperymentów w AWS. Wyniki pokazują, że zaproponowane rozwiązanie może znacząco zmniejszyć bariery wejścia do prowadzenia takich symulacji, czyniąc je bardziej praktycznymi w zastosowaniach i efektywnymi kosztowo. Wnioski z eksperymentów są uzupełnione o praktyczne wskazówki związane z prototypowaniem przy użyciu autorskiego klastra obliczeniowego opartego o komputery jednopłytkowe (Neural Simulation Cluster), oraz sugestie dotyczące dalszych prac badawczych i wdrożeń dla potoku symulacji neuronowej NSP.