Poster: GPU-Accelerated Artificial Neural Network for QSAR Modeling

Edward W. Lowe Jr., Nils Woetzel, Jens Meiler
Center for Structural Biology,
Vanderbilt University,
Nashville TN, USA
jens@jens-meiler.de

Keywords-Artificial Neural Network, opencl, gpu, QSAR, machine learning

I. INTRODUCTION

Machine learning techniques are widely used for the generation of models for the prediction of properties important in drug discovery such as absorption, distribution, metabolism, excretion, and toxicology (ADMET) properties[1]. Machine learning methods are commonly used in approximating non-linear separable data in QSAR studies[2]. These techniques have also been successfully applied to virtual high-throughput screening through the construction of QSAR models for the prediction of biological activity[3].

In this work, we have implemented an OpenCL artificial neural network (ANN) using back propagation of errors with added bias. ANNs have been applied to QSAR problems in previous work[4]. We demonstrate our implementation on biological data obtained through PubChem for 17β-hydroxysteroid dehydrogenase (HSD) inhibition, a target linked to neurodegeneration[5]. The data set contained 70k inactive compounds and 5k active compounds.

II. RESULTS

All C++ ANN trainings were performed on a Dell T3500 workstation equipped with 12GB RAM and an Intel(R) Xeon(R) W3570@3.20GHz running 64-bit CentOS 5.2. The OpenCL GPU ANN implementation trainings were performed on NVIDIA Tesla C1060, NVIDIA Fermi C2050, NVIDIA GTX 470, and ATI Radeon HD 5970.

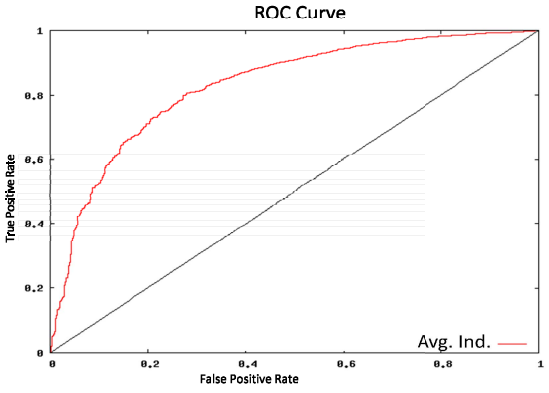

Figure 1. Average ROC curve from feature-optimized 5-fold cross-validated ANN training resulting in an average AUC of 0.83 and an enrichment of 5.9 for the independent data set.

Using an optimized feature set achieved through sequential forward feature selection, the ANN with a hidden layer of 8 neurons, eta of 0.001, and alpha of 0.5 achieved an average relative *rmsd* of 0.09 for the 5-fold cross validation trainings on the monitoring data set. The average enrichment achieved by the cross-validated trainings on the independent data set was 5.9 at a 1% cutoff while the average area under the curve was 0.83 (Fig 1). The OpenCL implementation achieves a speed-up up to 85-fold (Table I).

TABLE I
TRAINING TIMES IN MINUTES WITH SPEED-UP IN PARENTHESIS

Data Set	Xeon(R) @3.20GHz	Tesla C1060	Tesla C2050	GTX 470	ATI HD 5970
Full	5472	91.1(60)	67.2(81)	64.6(85)	118.2(47)

III. CONCLUSION

Here, we present a GPU-accelerated OpenCL implementation of a back-propagation artificial neural network for the creation of QSAR models for drug discovery and virtual high-throughput screening. A QSAR model for HSD achieved an enrichment of 5.9 and area under the curve of 0.83 on an independent data set which signifies sufficient predictive ability for virtual high-throughput screening efforts. The speed-up demonstrated on this data set allows for the complete cross-validated feature optimization of QSAR models based on ANNs within 24 hours on a workstation equipped with 4 consumer GPUs of $260 each (GTX 470), achieving performance equal to that of ~340 cores.

This GPU-accelerated ANN framework for the creation of optimized QSAR models from biological data will be available free of charge for academic users at http://www.meilerlab.org through a server interface.

REFERENCES

[1] T. Hou, J. Wang, W. Zhang, W. Wang, and X. Xu, "Recent advances in computational prediction of drug absorption and permeability in drug discovery," *Curr Med Chem,* vol. 13, pp. 2653-67, 2006. .

[2] Bleckmann and J. Meiler, "Epothilones: Quantitative Structure Activity Relations Studied by Support Vector Machines and Artificial Neural Networks," *QSAR Comb. Sci.,* vol. 22, pp. 719-721, 2003.

[3] M. Butkiewicz, R. Mueller, D. Selic, E. Dawson, and J. Meiler, "Application of Machine Learning Approaches on Quantitative Structure Activity Relationships," in *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, K. C. Wiese, Ed. Nashville, 2009.

[4] R. Mueller, A. L. Rodriguez, E. S. Dawson, M. Butkiewicz, T. T. Nguyen, S. Oleszkiewicz, A. Bleckmann, C. D. Weaver, C. W. Lindsley, P. J. Conn, and J. Meiler, "Identification of Metabotropic Glutamate Receptor Subtype 5 Potentiators Using Virtual High-Throughput Screening," *ACS Chemical Neuroscience,* vol. 1, pp. 288-305.

[5] Z. KriAtofikovA, M. BockovA, K. HegnerovA, A. BartoA, J. Klaschka, J. A?A?A?nA, D. A?A?povA, and J. Homola, "Enhanced levels of mitochondrial enzyme 17I?-hydroxysteroid dehydrogenase type 10 in patients with Alzheimer disease and multiple sclerosis," *Molecular BioSystems,* vol. 5, pp. 1174-1179, 2009.