Speeding up node label learning in unbalanced biomolecular networks through a parallel and sparse GPU-based Hopfield model

A. Petrini[2], M. Notaro[2], J. Gliozzo[1], G. Valentini[2], G. Grossi[2], M. Frasca[2]

[1] Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico, Università degli Studi di Milano
[2] AnacletoLab – Dipartimento di Informatica, Università degli Studi di Milano

Motivation

In network biology and medicine several problems can be modeled as node label inference in partially labeled networks. Nodes are biomedical entities (e.g. genes, patients) and connections represent a notion of functional similarity between entities. Usually, the class being predicted is represented through a labeling vector highly unbalanced towards negatives: that is only few positive instances (those associated with the class) are available. This fosters the adoption of imbalance-aware methodologies to accurately predict node labels. In addition, input data can be large-sized, since we may have millions of instances (e.g. in multi-species protein networks), thus requiring the design of efficient and scalable methodologies.

To address these problems, a parametric neural algorithm based on the Hopfield model, COSNet [1,2,3], has been proposed, leveraging the minimization of a Hopfield network energy through the usual sequential dynamics to achieve an asymptotically stable attractor representing a valuable prediction.

In this study, we propose a sparse and partially parallel implementation of COSNet, for sparse networks, which decomposes the input net in independent sets of neurons, each processed concurrently by hardware accelerators, like modern GPUs, while still keeping the overall dynamics sequential.

Methods

The Hopfield dynamics is decomposed in independent tasks by solving the graph coloring problem, that is assigning colors to the graph vertices so that adjacent vertices receive different colors. Thus, the units of the neural network are split into clusters of independent neurons, which are sequentially updated, whereas the single units within each cluster are updated simultaneously.

We simulate the algorithm on GPUs achieving a significant speed up with respect to the original sequential implementation and, at the same time, lowering memory requirements thanks to compressed memorization strategies, thus opening the possibility to face with prediction issues on big size instances. Also, a cooperative CPU multithreading – GPU model have been implemented, where the computations over different functional classes are carried independently by assigning each class to a different CPU thread.

Results

We tested both COSNet and COSNet-GPU on partially labeled networks containing genes belonging to *D. melanogaster* and *Homo sapiens* organisms for predicting respectively the Gene Ontology (GO) and the Human Phenotype Ontology (HPO) terms with 10-50 annotated genes. The algorithm behavior has been measured in terms of execution time and memory consumption.

Table 1 summarizes the results in term of speed-up and memory usage, when performing a 3-fold cross validation procedure. The results show significant reductions in both execution times and memory

consumption, and interestingly the improvement factors increases more than linearly with the number of nodes/genes. This also corroborates the fact that the proposed implementation nicely scales on big data.

Table 1. Comparison performances of sequential vs parallel version of COSNet.

Name	Nodes	Edges	GO/HPO Classes	Folds	COSNet time (s)	RAM (MB)	COSNet-GPU time (s)	RAM (MB)	GPU mem (MB)	Speed-up
Human	20257	2050328	1313	3	22816	13107	696	2600	65	32,4x
Fly	9361	825910	1568	3	5297	3555	259	1100	26	20,4x

References
[1] *A. Bertoni, M. Frasca, G. Valentini.* COSNet: a Cost Sensitive Neural Network for Semi-supervised Learning in Graphs*, Lecture Notes on Artificial Intelligence, vol. 6911, pp.219- 234, Springer. DOI*: 10.1007/978-3-642-23780-5_24.
[2] *M. Frasca, A. Bertoni, M. Re*, and *G. Valentini*, A neural network algorithm for semi-supervised node label learning from unbalanced data, *Neural Networks* 43, pp.84-98, 2013
[3] *M. Frasca, A. Bertoni, G. Valentini* UNIPred: Unbalance-aware Network Integration and Prediction of protein functions, *Journal of Computational Biology*, 22(12): 1057-1074, 2015