A GPU-Accelerated Skeleton Discovery for Gaussian Distribution Models

Motivation

The estimation of causal graphical models allows to solve important problems in many different domains. For example in genetics, gene regulatory networks can be seen as a practical embodiment of systems biology and can be used for drug design or diagnostics:

- Causal inference requires algorithmic support due to a rising complexity;
- Inefficiency of the common algorithms hinders its application in practice;
- Harness the processing power of the GPU to address the algorithm's computational complexity.

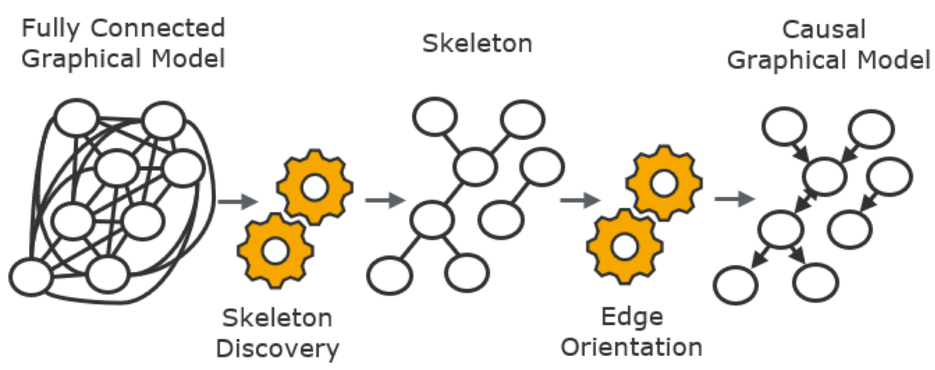

Figure: A schematic representation of the causal inference procedure

A Skeleton Discovery beyond GPU Memory Capacity

In previous work[3] a GPU-accelerated causal inference implementation has been proposed. While, it achieves significant speed-up, it is only applicable to datasets fitting into GPU memory. To overcome this limitation, we propose a block-wise skeleton discovery, shown in Algorithm 1. The algorithm splits a given dataset into smaller blocks, with the following characteristics:

- Blocks contain all data structures of their subpart
- Blocks' sizes are limited by the number of additional blocks required as separation sets for a given level
- Blocks expose enough work to occupy the compute resources of the GPU

Furthermore, the algorithm guarantees that all conditional independence tests are conducted, although data of required separation sets may be located in different blocks.

Causal Inference Procedure

In the recent years, the notion of causality has grown from a nebulous concept into a mathematical theory[1]. A conceptual algorithm for learning the causal graphical model operates in two phases[2]:

- *Skeleton Discovery,* use conditional independence tests to receive information about the underlying relationships.
- *Edge Orientation,* determine the orientation of the detected relationships to construct a causal graphical model.

Algorithm 1 Block-wise skeleton discovery of PC-stable algorithm
Input: Vertex set V, correlation matrix Cor
Output: Estimated skeleton $\mathcal{C}$, separation sets Sepset

```
Start with fully connected skeleton C and l = −1
repeat
    l = l + 1
    updateAdjacency(adj_{in,out})
    blocks = Split(Cor, adj_{in,out}, Sepset, l)
    for all b in blocks do
        Copy b to GPU
        if l == 0 then
            BlockwiseCITests(b)
        else
            sepsetblocks = SepSetCombinations(b, l, blocks)
            for all s in sepsetblocks do
                Copy s to GPU
                BlockwiseCITests(b, s)
            end for
        end if
        Copy b from GPU
    end for
    combineBlocks(blocks)
until each adjacent pair V_i, V_j in C satisfy |adj_out(V_i) \ {V_j}| ≥ l
return C, Sepset
```

Algorithm 1: Pseudocode of the block-wise skeleton discovery of PC-stable.

References:

1) J. Pearl. Causality: Models, Reasoning and Inference. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
2) P. Spirtes, C. N. Glymour, and R. Scheines. Causation, prediction, and search. MIT press, 2000.
3) C. Schmidt, J. Huegle, and M. Uflacker. Order-independent constraint-based causal structure learning for gaussian distribution models using GPUs. In Proceedings of the 30th International Conference on Scientific and Statistical Database Management (SSDBM '18). ACM, 2018, New York, NY, USA, Article 19, 10 pages.

Projektbeteiligte

Christopher Schmidt, Johannes Huegle and Matthias Uflacker
Hasso-Plattner-Institute
Enterprise Platform and Integration Concepts
August-Bebel-Str. 88
14482 Potsdam, Germany
{christopher.schmidt, johannes.huegle, matthias.uflacker}@hpi.de

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam