Towards a GPU-Accelerated Causal Inference

Motivation

The rise of the Internet of Things led to a growing interest into the analysis of massive complex datasets collected by sensors in industrial manufacturing processes:

- Structural knowledge allows to derive actionable insights;
- 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.

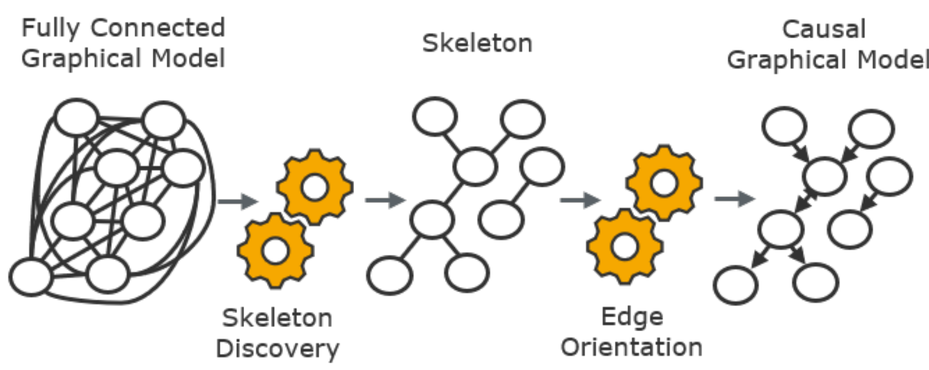

Figure: A schematic representation of the causal inference procedure

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.

Prelimanary Results

In our work, we developed an improved GPU-accelerated implementation for the calculation of the correlation matrix, called CUDA-shared. In our implementation, we reduce the number of accesses to global memory by utilizing the available shared memory of the GPU.
We evaluated our implementation based on real world gene expression[3] datasets and an industrial manufacturing dataset.
In the experiment we compare our implementation to an existing CUDA-enabled version from the R package gputools[4].
The empirical results, as depicted in the table on the right, show that the CUDA-shared version achieves a speed-up of up to a factor 1.5 .

	gputools	CUDA-shared
BR51	44.79	29.07
MCC	35.17	23.83
NCI-60	25.18	16.98
Saureus	184.13	124.89
Scerevisiae	564.83	357.46
Industrial	259.92	228.44

Table: Calculation of the correlation matrix gputools vs CUDA-shared - median of 100 executions in milliseconds

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) T. D. Le, T. Hoang, J. Li, L. Liu, and H. Liu. A fast PC algorithm for high dimensional causal discovery with multi-core pcs. CoRR, abs/1502.02454, 2015.
4) J. Buckner, M. Seligman, F. Meng, and J. Wilson. gputools: A Few GPU Enabled Functions, 2016.

Projektbeteiligte

Christopher Schmidt and Johannes Hügle
Hasso-Plattner-Institute
Enterprise Platform and Integration Concepts
August-Bebel-Str. 88
14482 Potsdam, Germany
Christopher.Schmidt@hpi.de
Johannes.Huegle@hpi.de

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam