

GPU-Accelerated Causal Structure Learning

Description

In this master's project, you develop a library that serves the data science community to enable causal structure learning (CSL) on top of latest GPU acceleration technologies.

CSL is a knowledge discovery technique to identify causal relationships and corresponding causal mechanisms from large quantities of data, for example, to understand the causes and effects of events in manufacturing processes under observation. Yet algorithms for CSL have high computational complexity and consequently are currently limited by their long runtimes. Existing work on GPU-accelerated CSL has shown potential for speed-up, but these works rely on manually written custom CUDA code which impedes extendibility and maintainability. You will close this gap by developing a library for CSL on GPUs.

To fulfill the project's goal, you will research existing state-of-the-art libraries and frameworks that support development on GPUs for data science (e.g., rapids, thrust, ...). Based on your research, you will decide on a suitable abstraction (through libraries or frameworks) for learning causal structures on GPUs. Using the abstraction, you will integrate methods that currently exist in manually written CUDA kernels into your package. In the end, we aim to make your contribution publicly available to the data science community.

Learning Goals

Through active participation in this project, you will:

- Improve your programming and teamwork skills
- Improve your research methodology and academic writing
- Gain practical experience in GPU acceleration for data science
- Understand methods of causal structure learning and deepen your knowledge on parallel programming

Skills

The core of the work is based upon existing CUDA-based implementations for GPU-accelerated constraint-based causal structure learning. Therefore, prior knowledge in one of the following areas is beneficial (C++, CUDA, Parallel Programming, GPU Programming) and understanding of the fundamentals of machine learning techniques (e.g., having attended the lecture "Causal Inference – Theory and Applications in Enterprise Computing", "Computational Statistics" or equivalent) is helpful.

Technology & Literature

- [1] <https://rapids.ai/> & <https://docs.rapids.ai/api/cudf/stable/>
- [2] <https://developer.nvidia.com/thrust>
- [3] Zarebavani, B.; Jafarinejad, F.; Hashemi, M.; Salehkaleybar, S.: cuPC: CUDA-Based parallel PC algorithm for causal structure learning on GPU. In IEEE Transactions on Parallel and Distributed Systems 31(03), mar 2020: pp. 530–542
- [4] Hagedorn, C.; Huegle, J.: GPU-Accelerated Constraint-Based Causal Structure Learning for Discrete Data. In Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2021, pp. 37–45.
- [5] Schmidt, C.; Huegle, J.; Uflacker, M.: 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, New York, NY, USA, 2018, pp. 19:1–19:10
- [6] Schmidt, C.; Huegle, J.; Horschig, S.; Uflacker, M.: Out-of-Core GPU-Accelerated Causal Structure Learning. In Algorithms and Architectures for Parallel Processing. Springer International Publishing, Cham, 2020, pp. 89–104
- [7] Spirtes, P.; Glymour, C.; Scheines, R.: Causation, Prediction, and Search, Second Edition. Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, USA, 2000
- [8] Pearl, Judea. "Causal inference in statistics: An overview." Statistics surveys 3 (2009): 96-146.

Organization

- Master's programs: ITSE, DE
- Extent: 8 SWS / 12 ECTS
- Content:
 - Group work (starting first week of lecture period; until end of semester)
 - Programming project
 - Final presentation (mid of February)
 - Research report (due: 31st of March)

Contact

You are welcome to contact us via email:

Christopher Hagedorn (Christopher.Hagedorn@hpi.de),

Johannes Huegle (Johannes.Huegle@hpi.de),

Dr. Rainer Schlosser (Rainer.Schlosser@hpi.de),

Dr. Michael Perscheid (Michael.Perscheid@hpi.de)