Recursive SQL and GPU-Support for In-Database Machine Learning

Maximilian E. Schüle[1]

Abstract: In machine learning, continuously retraining a model guarantees accurate predictions based on the latest data as training input. But to retrieve the latest data from a database, time-consuming extraction is necessary as database systems have rarely been used for operations such as matrix algebra and gradient descent. In this work, we demonstrate that SQL with recursive tables makes it possible to express a complete machine learning pipeline out of data preprocessing, model training and its validation. To facilitate the specification of loss functions, we extend the code-generating database system Umbra by an operator for automatic differentiation for use within recursive tables: With the loss function expressed in SQL as a lambda function, Umbra generates machine code for each partial derivative. We further use automatic differentiation for a dedicated gradient descent operator, which generates LLVM code to train a user-specified model on GPUs. We fine-tune GPU kernels at hardware level to allow a higher throughput and propose non-blocking synchronisation of multiple units. In our evaluation, automatic differentiation accelerated the runtime by the number of cached subexpressions compared to compiling each derivative separately. Our GPU kernels with independent models allowed maximal throughput even for small batch sizes, making machine learning pipelines within SQL more competitive.

References

[Sc22]   Schüle, M. E.; Lang, H.; Springer, M.; Kemper, A.; Neumann, T.; Günnemann, S.: Recursive SQL and GPU-support for in-database machine learning. Distributed Parallel Databases 40/2, pp. 205–259, 2022.

---

[1] University of Bamberg, maximilian.schuele@uni-bamberg.de