AutoGMap: Learning to Map Large-Scale Sparse Graphs on Memristive Crossbars

Bo Lyu, Shengbo Wang, Shiping Wen, Kaibo Shi, Yin Yang, Lingfang Zeng and Tingwen Huang

Abstract: The sparse representation of graphs has shown great potential for accelerating the computation of graph applications (e.g., social networks and knowledge graphs) on traditional computing architectures (CPU, GPU, or TPU). But, the exploration of large-scale sparse graph computing on processing-in-memory (PIM) platforms (typically with memristive crossbars) is still in its infancy. To implement the computation or storage of large-scale or batch graphs on memristive crossbars, a natural assumption is that a large-scale crossbar is demanded, but with low utilization. Some recent works question this assumption; to avoid the waste of storage and computational resource, the fixed-size or progressively scheduled "block partition" schemes are proposed. However, these methods are coarse-grained or static and are not effectively sparsity-aware. This work proposes the dynamic sparsity-aware mapping scheme generating method that models the problem with a sequential decision-making model, and optimizes it by reinforcement learning (RL) algorithm (REINFORCE). Our generating model long short-term memory (LSTM), combined with the dynamic-fill scheme generates remarkable mapping performance on the small-scale graph/matrix data (complete mapping costs 43% area of the original matrix) and two large-scale matrix data (costing 22.5% area on qh882 and 17.1% area on qh1484). Our method may be extended to sparse graph computing on other PIM architectures, not limited to the memristive device-based platforms.

Shiping Wen is the Director of the Intelligent Computing and Systems (ICS) Lab, Australian Artificial Intelligence Institute (AAII).

Publication: IEEE Transactions on Neural Networks and Learning Systems (https://ieeexplore.ieee.org/abstract/document/10103915)

DOI: 10.1109/TNNLS.2023.3265383