A Graph Theoretic Framework of Recomputation Algorithms for Memory-Efficient Backpropagation

Mitsuru Kusumoto, Takuya Inoue, Gentaro Watanabe, Takuya Akiba, Masanori Koyama

Recomputation algorithms collectively refer to a family of methods that aims to reduce the memory consumption of the backpropagation by selectively discarding the intermediate results of the forward propagation and recomputing the discarded results as needed. In this paper, we will propose a novel and efficient recomputation method that can be applied to a wider range of neural nets than previous methods. We use the language of graph theory to formalize the general recomputation problem of minimizing the computational overhead under a fixed memory budget constraint, and provide a dynamic programming solution to the problem. Our method can reduce the peak memory consumption on various benchmark networks by 36%~81%, which outperforms the reduction achieved by other methods.