NC STATE UNIVERSITY

PROJECT PROPOSAL FOR CSC791

Optimizing Sparse Tensor Operations for Efficient Data Analysis

Student:
Yuchao Su
200598905
ysu34@ncsu.edu

Instructor:
Prof. Jiajia Li

Department of Computer Science

October 18, 2024

1 Background

Sparse tensors are multi-dimensional arrays with a significant number of zero elements, used in fields like machine learning, healthcare analytics, and social network analysis. Unlike dense tensors, they require specialized storage formats—such as the Coordinate (COO) and Hierarchical Coordinate (HiCOO) formats—to efficiently handle memory and computation. Key operations include Tensor Element-Wise (TEW), Tensor-Times-Vector (TTV), and Tensor-Times-Matrix (TTM) products, each playing a critical role in algorithms for dimensionality reduction, decomposition, and data analysis. However, challenges such as irregular memory access patterns, parallelization complexity, and trade-offs between storage efficiency and computational speed must be addressed to achieve optimal performance on CPUs and GPUs. Understanding and benchmarking these operations helps unlock new insights into improving tensor-based applications across different hardware platforms.

2 Motivation

Sparse tensor operations are essential for efficiently handling multi-dimensional data in many real-world applications, including machine learning, healthcare, and scientific computing. However, implementing these operations is challenging due to irregular memory access patterns, parallelization issues, and trade-offs between storage efficiency and speed. This project offers an opportunity to build practical expertise in sparse tensor formats and operations, which will enhance the understanding needed to tackle these challenges. In addition to the academic value, this project aligns with my ongoing work on hardware support for sparse tensor operations. Gaining a fundamental understanding of coefficient tensor operations through this project will not only make these concepts more accessible for my research but also provide valuable inspiration for future work.

3 Description of Tasks

1. Implementation of Key Operations:
    - Tensor Element-Wise (TEW) Operations: Addition, subtraction, and multiplication between two tensors.

  - Tensor-Times-Vector (TTV) Operations: Multiplying tensors with vectors along a specific mode.

2. Use of Sparse Data Formats:
    - Implement COO and HiCOO formats to understand their impact on computational efficiency and storage.
    - Compare the performance between the formats on sample datasets.
3. Performance Evaluation:
    - Measure and analyze the performance of each operation on CPUs and GPUs(Optional).
    - Provide insights into how hardware architecture influence performance.
4. Documentation:
    - A comprehensive report detailing the implementation process, challenges encountered, and insights gained..
    - Code repository with instructions for running the implemented operations.
    - Poster with a brief project description and analysis of results

4 Time Plan

- Week1(Oct 18 - Oct 25)
    - Read and understand the paper[1], focusing on the HiCOO formats and basic sparse tensor operations.
    - Explore the required tools and libraries for implementation (Python with NumPy/SciPy or C++).
    - Plan datasets to be used for experiments.
- Week2(Oct 26 - Nov 1)
    - Implement TEW and TTV operations using COO formats.
- Week3(Nov 2 - Nov 8)

  - Extend TEW and TTV operations to the HiCOO format.
  - Conduct initial and basic tests to ensure correctness.
- Week4(Nov 9 - Nov 18)
  - Run extensive performance tests on CPUs with real and synthetic datasets.
  - Collect and analyze performance data, comparing COO and HiCOO formats.
- Nov 19 - Nov 21
  - Prepare the project poster, summarizing findings, challenges, and insights.
  - Finalize code and submit to the course platform.

5 Tools and Resources

- Language: C++
- Hardware: Plateform: ARC cluster.
- Reference: A Sparse Tensor Benchmark Suite for CPUs and GPUs[1]

References

[1] J. Li, M. Lakshminarasimhan, X. Wu, A. Li, C. Olschanowsky, and K. Barker, "A sparse tensor benchmark suite for cpus and gpus," in *2020 IEEE International Symposium on Workload Characterization (IISWC)*, 2020, pp. 193–204.