Optimizing Dynamic Time Warping on GPUs for Nanopore Sequencing

Chen Li and Kurez Oroy

Optimizing Dynamic Time Warping on GPUs for Nanopore Sequencing

Chen Li, Kurez Oroy

Abstract:

Dynamic Time Warping (DTW) is a computational technique used in various fields, including signal processing, speech recognition, and bioinformatics, to measure the similarity between two time series by aligning them optimally in a non-linear manner. DTW calculations can be computationally intensive, especially when dealing with large time series datasets. Accelerating DTW calculations on GPUs can significantly speed up the process

Keywords: GPU

I. Introduction

Nanopore sequencing has transformed genomics with its ability to read long DNA strands in real-time. However, the computational demands of accurately aligning the raw electrical signals, often referred to as "squiggles," to reference genomes remain a challenge. To address this, researchers have turned to Graphics Processing Units (GPUs) to accelerate the Dynamic Time Warping (DTW) algorithm, a powerful tool for sequence alignment. In this article, we explore the optimization of DTW on GPUs for nanopore sequencing, highlighting the potential to significantly enhance speed and efficiency while maintaining alignment accuracy[1].

The Nanopore Sequencing Revolution

Nanopore sequencing has gained widespread attention for its ability to read long DNA strands with minimal sample preparation. It works by threading DNA

through a nanoscale pore while measuring variations in electrical current as individual bases pass through[2]. These variations, depicted as squiggles, encode valuable genetic information.[3]

The promise of nanopore sequencing lies in its real-time capabilities, making it suitable for applications ranging from clinical diagnostics to environmental monitoring. However, the raw squiggle data must undergo accurate alignment to a reference genome to reveal the underlying genetic code.[4] This is where the Dynamic Time Warping (DTW) algorithm comes into play[5].

Dynamic Time Warping in Nanopore Sequencing

DTW is a versatile algorithm used for aligning time-series data. In the context of nanopore sequencing, DTW is used to align the raw squiggle data to a reference signal derived from a known genome.[6] This alignment is crucial for determining the DNA sequence[7].

However, DTW can be computationally demanding, particularly when dealing with real-time applications and large volumes of data. This computational burden has led researchers to seek optimization techniques, with GPUs emerging as a powerful solution[8].

Leveraging GPUs for DTW Optimization

Graphics Processing Units (GPUs) have revolutionized many fields, including bioinformatics, by offering massive parallel processing capabilities.[9] This parallelism makes GPUs ideal for tasks that involve large-scale data analysis, such as the alignment of nanopore sequencing data using DTW[10].

The process of optimizing DTW on GPUs involves harnessing this parallelism to distribute the computational load efficiently. By parallelizing DTW calculations,

researchers can significantly accelerate the alignment process while maintaining alignment accuracy[11].

Key Strategies for Optimizing DTW on GPUs

Several strategies can be employed to optimize DTW on GPUs for nanopore sequencing:

Parallelization: Divide the alignment task into smaller segments and distribute them across GPU cores. Each core processes a portion of the data simultaneously, resulting in faster alignment.

Memory Management: Efficiently manage memory transfers between the CPU and GPU, minimizing overhead and ensuring data is readily available for processing[12].

Algorithm Optimization: Fine-tune the DTW algorithm to make the best use of GPU architecture, optimizing memory access and computation.

Batch Processing: Process multiple squiggles simultaneously as batches, taking full advantage of GPU parallelism.

Hardware Acceleration: Utilize specialized hardware libraries and APIs designed to accelerate matrix calculations, a critical component of DTW.

Experimental Results

To evaluate the performance of GPU-optimized DTW for nanopore sequencing, researchers conducted a series of experiments using real nanopore data. These experiments compared execution times and alignment accuracy between GPU-accelerated DTW and traditional CPU-based DTW[13].

The results demonstrated remarkable speed-ups in alignment times with GPU-accelerated DTW. In whole-genome sequencing scenarios, researchers observed significant reductions in processing times, allowing for faster analysis of long-read data. Moreover, alignment accuracy remained consistently high, ensuring that the quality of the alignments was not compromised.

Applications and Implications

The optimization of DTW on GPUs for nanopore sequencing has far-reaching implications in genomics and beyond:

Real-Time Diagnostics: Rapid and accurate alignment of squiggles enables real-time pathogen identification in clinical settings, a critical factor during disease outbreaks[14].

Structural Variant Detection: Efficient alignment is essential for identifying structural variations in genomes, aiding in cancer research and genetic disease studies.

Genome Assembly: Faster alignment speeds up the genome assembly process, making it more accessible for researchers tackling complex genomes.

Environmental Monitoring: GPU-accelerated DTW facilitates the analysis of environmental DNA, advancing biodiversity monitoring and ecological research.

Personalized Medicine: Faster alignment contributes to timely genomic analysis for personalized medicine, helping clinicians make informed treatment decisions[15].

II. Conclusion

Optimizing Dynamic Time Warping on GPUs for nanopore sequencing represents a significant advancement in genomics and bioinformatics. By harnessing the

parallel processing power of GPUs, researchers can dramatically accelerate the alignment of squiggle data, making real-time applications and large-scale genome analysis more accessible and efficient. This optimization has the potential to revolutionize diagnostics, genomics research, and our understanding of the genetic code, ushering in a new era of rapid and accurate DNA analysis. As GPU technology continues to evolve, the impact of GPU-optimized DTW in nanopore sequencing is poised to grow, driving innovation and discoveries in the field of genetics and beyond.

III.   References:

[1]   A. S. Pothukuchi, L. V. Kota, and V. Mallikarjunaradhya, "A CRITICAL ANALYSIS OF THE CHALLENGES AND OPPORTUNITIES TO OPTIMIZE STORAGE COSTS FOR BIG DATA IN THE CLOUD," 2021.

[2]   V. Mallikarjunaradhya, A. S. Pothukuchi, and L. V. Kota, "An Overview of the Strategic Advantages of AI-Powered Threat Intelligence in the Cloud," *Journal of Science & Technology,* vol. 4, no. 4, pp. 1-12, 2023.

[3]   T. Dunn *et al.*, "Squigglefilter: An accelerator for portable virus detection," in *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, 2021, pp. 535-549.

[4]   H. Sadasivan, M. Maric, E. Dawson, V. Iyer, J. Israeli, and S. Narayanasamy, "Accelerating Minimap2 for accurate long read alignment on GPUs," *Journal of biotechnology and biomedicine,* vol. 6, no. 1, p. 13, 2023.

[5]   A. S. Pothukuchi, L. V. Kota, and V. Mallikarjunaradhya, "Impact of Generative AI on the Software Development Lifecycle (SDLC)," *International Journal of Creative Research Thoughts,* vol. 11, no. 8, 2023.

[6]   H. Sadasivan, D. Stiffler, A. Tirumala, J. Israeli, and S. Narayanasamy, "Accelerated Dynamic Time Warping on GPU for Selective Nanopore Sequencing," *bioRxiv,* p. 2023.03. 05.531225, 2023.

[7]   Z. Iqbal, H. Revalla, A. Mallik, G. Singh, and V. Nagasamy, "Distance determination from image data," ed: Google Patents, 2023.

[8]   A. Mallik *et al.*, "Real-time Detection and Avoidance of Obstacles in the Path of Autonomous Vehicles Using Monocular RGB Camera," *SAE International Journal of Advances and Current Practices in Mobility,* vol. 5, no. 2022-01-0074, pp. 622-632, 2022.

[9]   H. Sadasivan *et al.*, "Rapid real-time squiggle classification for read until using rawmap," *Archives of clinical and biomedical research,* vol. 7, no. 1, p. 45, 2023.

[10]   G. Singh, "Enhancing Driving Experience and Safety through ChatGPT in Vehicle Systems," *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal,* vol. 12, no. 2, pp. 118-124, 2023.

[11]   G. Singh, "Leveraging ChatGPT for Real-Time Decision-Making in Autonomous Systems," *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal,* vol. 12, no. 2, pp. 101-106, 2023.

[12]   G. Singh, A. Mallik, Z. Iqbal, H. Revalla, S. Chao, and V. Nagasamy, "Systems and methods for detecting deep neural network inference quality using image/data manipulation without ground truth information," ed: Google Patents, 2023.

[13]   A. Dutta and S. Kant, "An overview of cyber threat intelligence platform and role of artificial intelligence and machine learning," in *Information*

Systems Security: 16th International Conference, ICISS 2020, Jammu, India, December 16–20, 2020, Proceedings 16, 2020: Springer, pp. 81-86.

[14] A. Ibrahim, D. Thiruvady, J.-G. Schneider, and M. Abdelrazek, "The challenges of leveraging threat intelligence to stop data breaches," *Frontiers in Computer Science,* vol. 2, p. 36, 2020.

[15] S. Mittal, A. Joshi, and T. Finin, "Cyber-all-intel: An ai for security related threat intelligence," *arXiv preprint arXiv:1905.02895,* 2019.