EasyChair Preprint

№ 10937

Real-Time Squiggle Classification for Read Until

Chris Liu and Kurez Oroy

September 21, 2023

Real-Time Squiggle Classification for Read Until

Chris Liu, Kurez Oroy

Abstract:

Nanopore sequencing has introduced the concept of "Read Until," enabling selective sequencing of specific DNA regions of interest. Real-time squiggle classification plays a pivotal role in Read Until by rapidly identifying target sequences. This article explores the development of real-time squiggle classification methods to enhance the performance of Read Until. By accurately classifying squiggles as they are generated during sequencing, researchers can make precise decisions about which reads to continue, enabling efficient targeted sequencing.

I. Introduction:

In the realm of genomics, technological advancements have continually driven progress, allowing us to explore the intricacies of DNA and RNA with ever-increasing precision and efficiency. One such groundbreaking innovation is Nanopore Sequencing, a transformative approach that has redefined the landscape of DNA sequencing. In this comprehensive exploration, we delve into the principles, applications, advantages, and challenges of Nanopore Sequencing, uncovering its pivotal role in genomics research and diagnostics.At the heart of Nanopore Sequencing lies a simple yet ingenious concept: threading a single strand of DNA or RNA through a nanometer-sized pore (the "nanopore") and monitoring the changes in electrical current as the molecule traverses the pore. These changes in current are characteristic of the DNA bases (adenine, thymine, cytosine, and guanine) and can be used to decipher the sequence.

There are two primary types of nanopores used in sequencing:

- **Biological Nanopores**: These are naturally occurring protein channels, such as α-hemolysin, used in early nanopore sequencing methods.
- **Solid-State Nanopores**: Synthetic nanopores created in solid-state materials, like silicon nitride, offer more control and scalability.

2. **Readout Technologies**:

- **Ion Current Measurement**: The most common method, where changes in ionic current are detected as the DNA molecule passes through the nanopore.
- **Tunneling Current Measurement**: This approach measures the electronic tunneling current across the pore.

Nanopore sequencing has revolutionized genomics with its real-time, long-read capabilities. One of the most promising features of nanopore sequencing is "Read Until," a concept that allows selective sequencing of specific DNA regions of interest. Real-time squiggle classification is at the heart of Read Until, as it determines which reads are relevant to the targeted regions. This article delves into the importance of real-time squiggle classification and its applications in genomics.

II. The Significance of Real-Time Squiggle Classification

Real-time squiggle classification refers to the process of rapidly categorizing the raw electrical signals, known as squiggles, generated during nanopore sequencing. These squiggles contain valuable genetic information, and accurate classification is essential for identifying target sequences in real-time[1]. The significance of real-time squiggle classification includes:

Efficient Targeted Sequencing: Real-time classification allows for the efficient selection of reads that align with specific genomic regions of interest, reducing sequencing time and costs.[2]

Rapid Pathogen Detection: In clinical diagnostics, real-time squiggle classification can swiftly identify pathogenic DNA, enabling rapid disease diagnosis.[3]

Genomic Analysis: Researchers can focus on specific genes or regions of interest for in-depth genomic analysis, speeding up discoveries.[4]

Personalized Genomics: Real-time squiggle classification contributes to personalized genomics by targeting relevant genetic variants for individualized medicine.[5]

Challenges in Real-Time Squiggle Classification

Real-time squiggle classification presents several challenges:

Speed: Squiggle classification must occur rapidly to enable Read Until's real-time capabilities.[6]

Accuracy: High accuracy is crucial to prevent false positives or negatives when identifying target sequences.[7]

Data Volume: Nanopore sequencers generate large volumes of data, and real-time classification must keep up with this data throughput.[8]

Methods for Real-Time Squiggle Classification

Several methods and algorithms have been developed for real-time squiggle classification:

Machine Learning: Supervised machine learning models can be trained to classify squiggles based on features that distinguish target sequences from non-target sequences.

Signal Processing: Signal processing techniques can extract relevant features from squiggles and classify them based on patterns or characteristics.

Hybrid Approaches: Combining machine learning and signal processing methods can enhance classification accuracy and speed.[9]

III. Conclusion:

Accelerating real-time squiggle classification using GPUs and parallel processing techniques offers a transformative solution for nanopore sequencing. By significantly boosting classification speed, researchers can harness the full potential of real-time sequencing, enhancing efficiency, and enabling rapid decisions in genomics applications. The continued development of GPU-accelerated squiggle classification methods holds the key to unlocking new possibilities in clinical diagnostics, environmental monitoring, genomics research, and beyond.[10]

IV. References:

[1] 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.

[2] 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.

[3] H. Sadasivan, "Accelerated Systems for Portable DNA Sequencing," University of Michigan, 2023.

[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] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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