Quantum Pattern Recognition for Local Sequence Alignment

Konstantinos Prousalis, Nikos Konofaos

Presented by Spencer King, Mingi Ryu

Qiskit Summer Jam

July 1, 2020

Background

- Pattern recognition is a key problem in computational biology
- Sequence alignment: finding regions of similarity between objects
  - Typically used with protein, DNA, or RNA
  - Commonly used in other fields as well (e.g. NLP and finance)
- Global vs Local alignment
  - Global: sequences must be exact matches
  - Local: Interested in longest common substring
- Common classical algorithms for local sequence alignment:
  - Smith-Waterman
  - Needleman-Wunsch

Smith-Waterman Algorithm

- Create a scoring matrix H with size $(n + 1) * (m + 1)$ where n and m are the lengths of the sequences to compare
- Initialize the first row and column of H to 0
- Populate the scoring matrix:

$$H_{ij} = \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j), \\ \max_{k \geq 1} \{H_{i-k,j} - W_k\}, \\ \max_{\ell \geq 1} \{H_{i,j-\ell} - W_l\}, \\ 0 \end{cases}$$

where $1 \leq i \leq n$ and $1 \leq j \leq m$.

- The best alignment is determined by following the path from the highest score to a score of 0

Smith-Waterman Algorithm

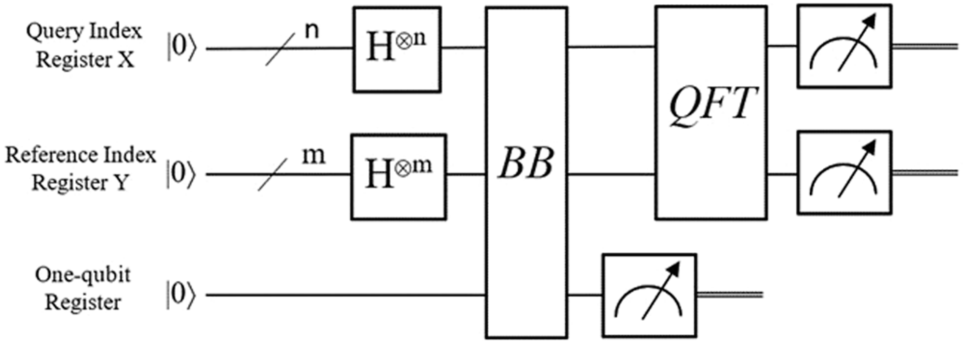
		C	A	C	C	G	T	A	A
	0	0	0	0	0	0	0	0	0
A	0	0	3	1	0	0	0	3	3
A	0	0	3	1	0	0	0	3	6
C	0	3	1	6	4	2	0	1	4
C	0	3	1	4	9	7	5	3	2
A	0	1	6	4	7	6	4	8	6
G	0	0	4	3	5	10	8	6	5
T	0	0	2	1	3	8	13	11	9
C	0	3	1	5	4	6	11	10	8
G	0	1	0	3	2	7	9	8	7

3	6	9	7	10	13
A	C	C	-	G	T
A	C	C	A	G	T

Smith-Waterman Algorithm - Complexity

- Cubic in time
- Quadratic in space
- Many algorithms have been proposed to handle larger problems
  - Computationally more efficient, but with significantly reduced generality
- Quantum algorithms could potentially yield significant reductions in both time and space

Quantum Algorithm



- Input data preparation
  - Use the Smith-Waterman algorithm to construct a substitution matrix.
- Black-Box run
  - Initialize the first two registers to n-qubit and m-qubit $|0>$ states and apply the Hadamard gate. The third register is a one qubit $|0>$ state and is used for the output function $f(x, y)$.
- Superposition by measurement
  - Apply a measurement to the third register to verify the superposition state.
- QFT application
  - Apply the Quantum Fourier Transform to transform the superposition state such that the peak of the wavefunction can be used as a way of pattern detection (feature selection).
- Analysis of pattern localization
  - Once we have significant pattern detection, we can make use of classical computation to localize the pattern length and direction.

Superposition by Measurement

- If $f = 1$ is measured
  - $|\Psi>$ is prepared as a superposition of $|x>$ and $|y>$ of all points.
- If $f \neq 1$
  - Assume an ideal black box was adopted.
- When $f = 1$ the superposition is prepared as:

$$|\Psi\rangle = \frac{1}{\sqrt{\rho S}} \sum_{\ell=1}^{\rho S} |z_\ell\rangle$$

- We feel there is too much going on in this step to fully discuss it here. We recommend reviewing the paper for more information.

QFT Application

- QFT is applied to the basis $|z>$ and transforms the superposition state:

$$QFT|\Psi\rangle = \sum_{k=0}^{S-1} \sum_{\ell=1}^{\rho S} \frac{1}{S\sqrt{\rho}} \exp(2\pi i \frac{z_\ell k}{S})|k\rangle$$

- Measuring $|k>$ may show peaks of factor $|z>$ at certain k values, giving insight into potential regions of similarity.
  - k will simply yield noise for non-zero values if there is no region of similarity

Issues

- Transforming matrix input into linear array.
  - The paper proposes adjusting the angle of the spatial light modulator to avoid needing to transform the input matrix. We do not have access to a spatial light modulator and thus needed to transform the matrix classically. Unfortunately, the authors do not go into detail about how to accomplish this and we are unsure if our approach is correct.
- Preparing n-qubit and m-qubit strings based on our matrix
  - The authors do not elaborate on this process. We believe the process would be apparent if we had stronger quantum backgrounds.
- General lack of quantum knowledge
  - Much of what comes later in the paper, for example measuring the superposition state $|\Psi>$ and applying the QFT was beyond our current knowledge. We feel that we need stronger quantum backgrounds in order to be effective, at least in this domain.

Future

- Study
  - The paper made it clear to us there is a significant amount of knowledge we lack. In order to pursue this further we will need to study quite a bit.

- Build off of existing Qiskit tutorials
  - There are Qiskit tutorials for doing simple string comparison and bit string comparison. Essentially, we could use these for global sequence alignment.
  - We could explore encoding longer sequences such that they can be compared with a limited number of qubits, or we could try to adapt this into a more beginner-friendly local alignment tutorial.

Resources

- GitHub repository
  - QiskitSummerJam-LocalSequenceAlignment
- Relevant Papers
  - Quantum Pattern Recognition for Local Sequence Alignment
  - A Quantum Pattern Recognition Method for Improving Pairwise Sequence Alignment